BMC Medical Informatics and Decision Making

RESEARCH

Open Access

A comprehensive framework to estimate the frequency, duration, and risk factors for diagnostic delays using bootstrapping-based simulation methods

Aaron C Miller[1*], Joseph E Cavanaugh[2], Alan T Arakkal[2], Scott H Koeneman[2] and Philip M Polgreen[1]

Abstract

Background The incidence of diagnostic delays is unknown for many diseases and specific healthcare settings. Many existing methods to identify diagnostic delays are resource intensive or difficult to apply to different diseases or settings. Administrative and other real-world data sources may offer the ability to better identify and study diagnostic delays for a range of diseases.

Methods We propose a comprehensive framework to estimate the frequency of missed diagnostic opportunities for a given disease using real-world longitudinal data sources. We provide a conceptual model of the disease-diagnostic, data-generating process. We then propose a bootstrapping method to estimate measures of the frequency of missed diagnostic opportunities and duration of delays. This approach identifies diagnostic opportunities based on signs and symptoms occurring prior to an initial diagnosis, while accounting for expected patterns of healthcare that may appear as coincidental symptoms. Three different bootstrapping algorithms are described along with estimation procedures to implement the resampling. Finally, we apply our approach to the diseases of tuberculosis, acute myocardial infarction, and stroke to estimate the frequency and duration of diagnostic delays for these diseases.

Results Using the IBM MarketScan Research databases from 2001 to 2017, we identified 2,073 cases of tuberculosis, 359,625 cases of AMI, and 367,768 cases of stroke. Depending on the simulation approach that was used, we estimated that 6.9–8.3% of patients with stroke, 16.0-21.3% of patients with AMI and 63.9–82.3% of patients with tuberculosis experienced a missed diagnostic opportunity. Similarly, we estimated that, on average, diagnostic delays lasted 6.7–7.6 days for stroke, 6.7–8.2 days for AMI, and 34.3–44.5 days for tuberculosis. Estimates for each of these measures was consistent with prior literature; however, specific estimates varied across the different simulation algorithms considered.

Conclusions Our approach can be easily applied to study diagnostic delays using longitudinal administrative data sources. Moreover, this general approach can be customized to fit a range of diseases to account for specific clinical

*Correspondence:
Aaron C Miller
miller@uiowa.edu

Full list of author information is available at the end of the article

characteristics of a given disease. We summarize how the choice of simulation algorithm may impact the resulting estimates and provide guidance on the statistical considerations for applying our approach to future studies.

Keywords Diagnosis, Diagnostic errors, Delayed diagnosis, Epidemiologic methods, Tuberculosis, Acute myocardial infarction, Stroke

Background

Diagnostic errors are a major contributor to morbidity, mortality and excess healthcare costs [1, 2]. Diagnostic delays are a common type of diagnostic error. For many diseases, timely diagnosis is essential for effective treatment, and for some diseases even minimal delays may significantly increase risk of patient harm [3, 4]. Identifying cases where diagnostic delays have occurred is a critical first step in studying the causes and consequences of diagnostic delays and for developing interventions to prevent delays. However, for many diseases and settings, the incidence of diagnostic delays is unknown or challenging to estimate [5, 6].

Historically, several approaches have been used to study diagnostic delays; these include retrospective chart reviews, autopsy studies, analysis of malpractice claims, and patient or clinician surveys [5, 7]. These approaches are highly informative, but have a number of limitations. For example, chart reviews are labor intensive, and have been primarily focused on single hospitals or health systems, thus limiting their generalizability. Other approaches, such as autopsy studies or malpractice claims may only apply to the most serious cases or diseases. Moreover, many approaches to study diagnostic delays have exclusively focused on hospital records or emergency department settings [8–13]. Yet, many opportunities to diagnose a disease occur in outpatient clinics,[14, 15] and patient care often occurs across a wide spectrum of disconnected facilities. Thus, longitudinal information spanning a wide variety of healthcare settings and covering a broad patient population is required to fully capture the diverse spectrum of diagnostic delays.

Another limitation of most investigations of diagnostic delays is that the criteria used to define a diagnostic delay must be specified *a priori*. Typically, expert evaluation must be used to determine the criteria to define a diagnostic delay based on what is known about the natural history of the disease prior to diagnosis. These criteria include validating the index diagnosis, describing the clinical signs and symptoms that indicate the disease was present prior to diagnosis, identifying the types of clinic records (e.g., notes, lab results, diagnostic codes, etc.) necessary to capture signs and symptoms of the disease, and selecting the biologically plausible period of time prior to the index diagnosis where an earlier diagnosis could have occurred [8, 9]. However, if a significant number of diagnostic delays occur among patients with atypical presentation or outside the time period evaluated, such cases may not be considered.

Another limitation with specifying criteria for a diagnostic delay in an *a priori* fashion is that some patients may meet criteria defining a delay simply by coincidence, especially if the criteria include common clinical signs or symptoms. For example, patients with tuberculosis may have a history of a cough prior to developing tuberculosis or patients may suffer from back pain prior to developing a spinal abscess. In such cases, symptoms may appear to be attributable to the disease, but are unrelated. Numerous investigations have relied on algorithms to identify diagnostic delays based on commonly occurring symptom criteria such as cough, fever, pain, headaches, malaise and fatigue, [8, 9, 12, 16] yet only a few have attempted to account for a coincidental or expected occurrence of such symptoms [17–19].

A growing number of investigators have begun to use longitudinal administrative and EMR-based data to identify diagnostic delays [11, 20–22]. These data allow inpatient, outpatient or emergency department (ED) records to be used in a retrospective approach, where evidence of a disease (e.g., symptom codes) is identified in visits prior to the definitive diagnosis. For example, visits associated with dizziness may be identified prior to a stroke diagnosis,[9, 13] or cough and fever may be identified prior to a tuberculosis diagnosis [16, 18]. Such visits are then considered potential missed opportunities if they occur during a specified *diagnostic opportunity window* - the time before the initial diagnosis where clinical disease may be present and where a diagnostic delay may occur (e.g., 10-days prior to a stroke diagnosis or 90-days prior to a tuberculosis diagnosis). This approach has been used to study a variety of diseases, including acute myocardial infarctions, strokes, subarachnoid hemorrhages, abdominal aortic aneurysms and tuberculosis,[8–13, 16, 17] and was formalized by Lieberman et al [20].

However, three methodological limitations exist with many of the current approaches to study diagnostic delays using observational data. First, as noted above, some signs and symptoms of disease observed prior to diagnosis might not represent actual diagnostic delays, but rather coincidental events that occur prior to the index diagnosis. Second, applications typically require investigators to pre-specify the diagnostic opportunity window prior to diagnosis when delays would be expected to occur. A window that is too long will tend to overestimate the number of diagnostic delays, while

a window too short will lead to underestimates. Third, diagnostic codes for symptoms associated with a diagnosis may be underutilized [23]. For example, patients with a cough may not receive a diagnostic code for cough and instead be assigned a code for pneumonia or respiratory infection. Relying solely on symptom-based codes would miss these visits, especially with conditions that may first be misdiagnosed as an alternative disease (e.g., pneumonia, asthma, or COPD instead of tuberculosis).

The purpose of this paper is to expand upon the existing literature using longitudinal observational data sources to study diagnostic delays, while providing a methodological framework to address the limitations highlighted above. Specifically, we describe a bootstrapping-based approach to estimate the number of "likely" missed diagnostic opportunities that individual patients experience and the duration of diagnostic delays. We also provide three different simulation algorithms that may be used to implement this bootstrapping approach. In addition, we describe possible methods for estimating the frequency of diagnostic delays in conjunction with the bootstrapping technique, considerations for choosing an estimation procedure, and a statistical software package, each of which allow these methods to be customized to a wide range of diseases. This work expands upon the basic conceptual framework described by Liberman et al.[20]. It also builds upon the methodological approach utilized by Waxman et al.[17] to separate observed and expected trends in symptomatic visits prior to diagnosis. Moreover, this study generalizes the methods the study authors have previously developed to investigate diagnostic delays associated with a range of diseases [18, 19, 24, 25].

The following sections summarize our approach along with three empirical applications and is organized as follows. We start by describing the conceptual framework behind our approach. Next, we outline the basic simulation framework along with three algorithms that may be considered to resample the data and implement our bootstrapping approach. We then describe some of the estimation procedures that may be used to determine the parameters necessary to implement the simulation. We also describe sensitivity analyses that may be considered. We then apply our method to three diseases where diagnostic delays have been previously investigated using large administrative data sources – tuberculosis, acute myocardial infarction (AMI), and stroke. Finally, we describe how results for this disease may differ across the different simulation algorithms and estimation procedures. We conclude by discussing considerations for future investigations.

Methods

Theoretical and Conceptual Framework

We define a missed diagnostic opportunity as a healthcare encounter where signs or symptoms of a disease are present, but where the diagnosis is not made or an incorrect diagnosis is applied. Our methodological framework is based on the following fundamental assumption: *for a given disease, a portion of patients will experience missed diagnostic opportunities prior to the index diagnosis of the disease, and such missed opportunities will be reflected by a greater than expected number of healthcare visits where signs and symptoms of the disease are present.*

To identify potential missed opportunities, we start by computing the number of visits prior to the index diagnosis where signs and symptoms of the disease of interest are present. We expand upon the *symptom-disease pair* concept from Lieberman, et al. [20] to include what we term as "symptomatically similar diagnosis-disease pairs" where a *symptomatically similar diagnosis* (SSD) encompasses not only signs and symptoms or related diagnoses, but also tests or procedures that may suggest the presence of the disease [20]. SSD-related visits may be identified using diagnosis codes (e.g., ICD-9-CM/ICD-10-CM), procedure codes (CPT or ICD), medication claims, or other structured data elements. We generally categorize SSDs into one of three types of events (and this list may be expanded upon based on expert-feedback or biological plausibility):

1. **General symptoms** of the disease of interest (e.g., cough, fever, weight loss, or hemoptysis before tuberculosis).
2. **Symptomatically-similar diseases or syndromes** that share similar symptoms to the disease of interest and subsequently may be confused for the disease of interest (e.g., diagnoses of pneumonia, influenza, or bronchitis before tuberculosis).
3. **Testing, imaging, physical-exam-based diagnoses, or treatments** that are associated with symptoms of the disease of interest (e.g., infection testing, chest x-rays, diagnoses of anemia or swollen lymph nodes before tuberculosis).

We analyze diagnostic opportunities by evaluating the trend in SSD-related visits prior to the index diagnosis. Figure 1 depicts SSD-related visits prior to the index tuberculosis diagnosis for 2,073 patients (described below). The x-axis depicts the number of days prior to the index tuberculosis diagnosis, and the y-axis depicts the number of visits that occurred on a given day that had an SSD-related diagnosis. From Fig. 1, we see that there is a large visible spike in the number of visits for SSD conditions that might be related to tuberculosis in the time just prior to the initial diagnosis. The dramatic increase appears to occur around 90–100 days prior to the index diagnosis, a time period consistent with prior

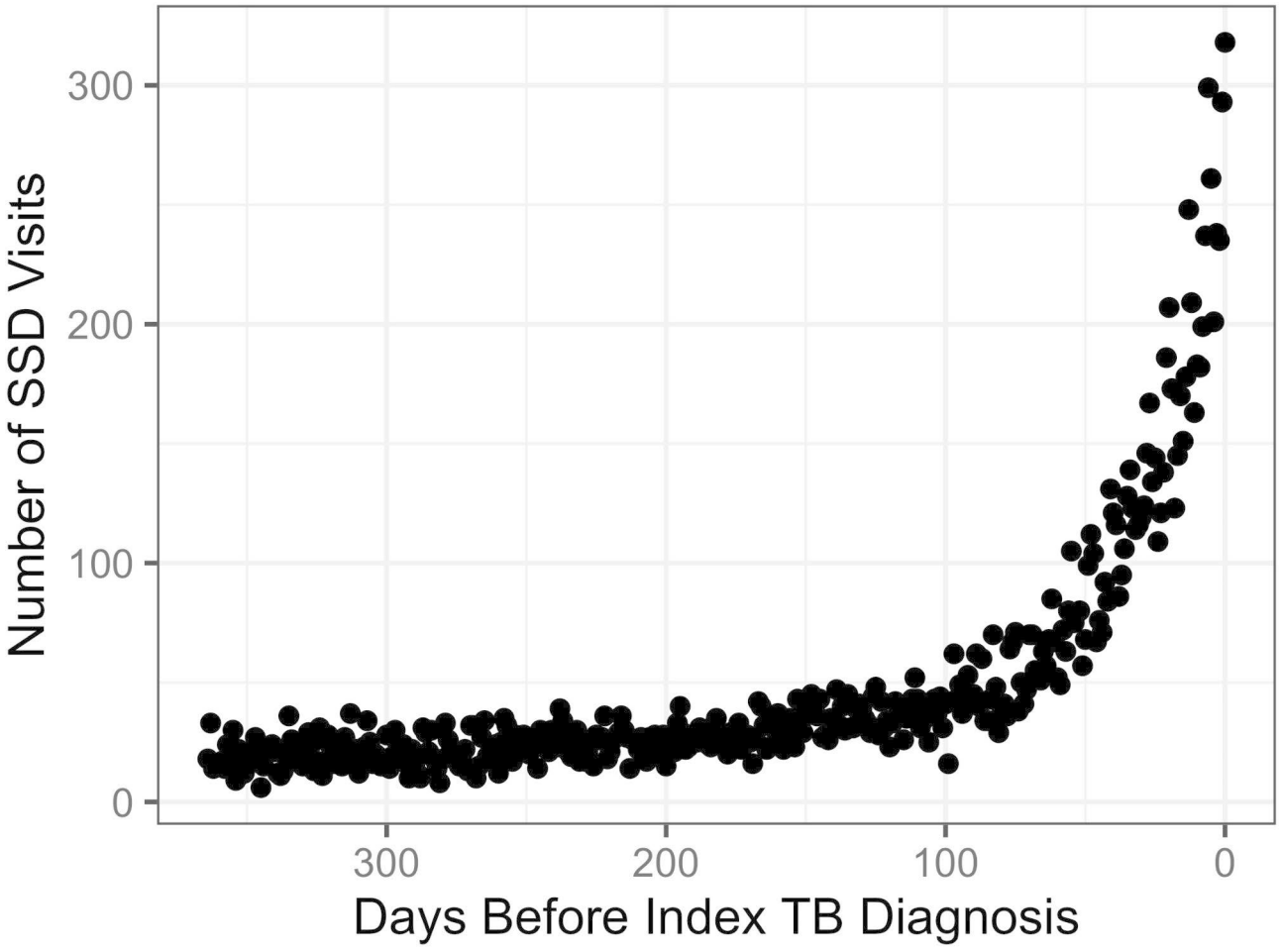

Fig. 1 Count of SSD-related visits prior to index tuberculosis diagnosis aggregated across all patients with tuberculosis diagnosis

investigations for when diagnostic delays for tuberculosis might occur [16, 26].

The trends presented in Fig. 1 have been broadly identified for a wide range of diseases and numerous studies have used this spike in healthcare utilization prior to diagnosis as a marker for diagnostic opportunities [8, 9, 17–19, 24, 25]. We assume this trend depicts two periods of activity: (1) a window just prior to diagnosis where SSD-related visits appear to dramatically increase, which we refer to as the *diagnostic opportunity window* and (2) a period further before diagnosis where SSD visits appear to exhibit either a stable or slightly increasing trend. These two periods are highlighted in Fig. 2, separated by the dashed-grey line and represent points in time where diagnostic delays are likely to occur (to the right) or unlikely to occur (to the left). The period prior to the diagnostic opportunity window is depicted by a relatively gradual upward trend in SSD visits. This period may capture risk factors for this disease or the natural history of the disease, but it is generally assumed to not reflect a significant portion of missed diagnostic opportunities. Note, Supplementary Methods S1 describe several factors that may drive this upward trend along with associated considerations that arise when modeling this period (e.g., both linear and non-linear trends may be used).

We build upon the distinction between observed and expected trends outlined by Waxman et al. (2018) and our prior empirical work[18, 19] to distinguish missed opportunities from coincidental care. Specifically, the red line in Fig. 2, depicts the baseline *expected* number of SSD visits – these represent SSD-related visits that are expected to occur in the absence of diagnostic delays. The solid-red line, reflects the observed trend in SSD visits prior to the *diagnostic opportunity window*; the dashed-red line reflects this trend extrapolated to the diagnostic opportunity window, as the number of SSD visits one would expect if the disease of interest was absent. This extrapolation reflects a type of *case-crossover design*, where the period prior to diagnostic opportunity window is used as a control period to estimate the expected number of visits (if the disease were absent). The visits approximated by the red shaded area, below the *expected* trend line, represent the number of SSD

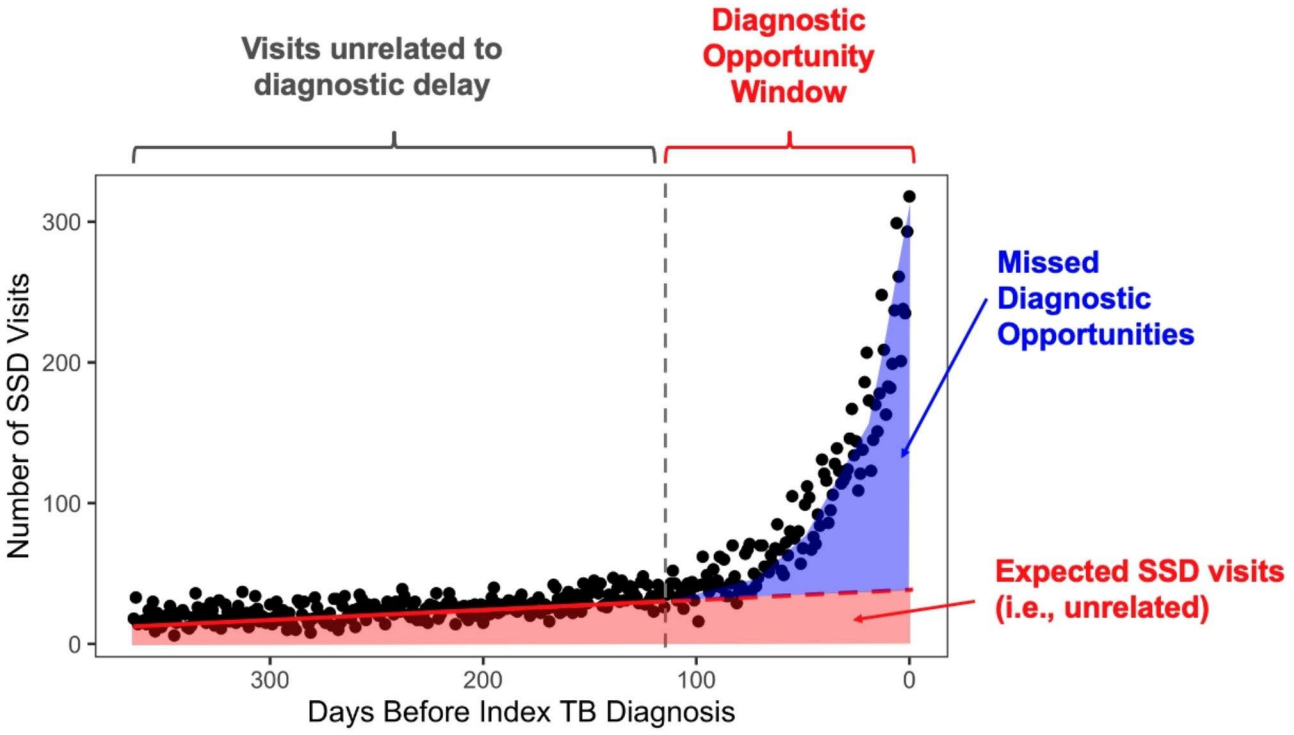

Fig. 2 Diagram of conceptual framework representing the number of missed diagnostic opportunities. The diagnostic opportunity window represents the period of time where diagnostic opportunities may occur. The red line depicts the trend in the number of SSD visits that would be expected to occur in absence of diagnostic delays. The blue curve represents the observed trend in SSD visits during the diagnostic opportunity window. The shaded blue region corresponds to the number of missed diagnostic opportunities. The shaded red region corresponds to the number of expected SSD visits

visits that would be expected to occur in the absence of diagnostic opportunities.

The blue curve in Fig. 2 represents the *observed* trend in SSD visits during the diagnostic opportunity window. The shaded blue area, between the observed and expected trends inside the diagnostic opportunity window, represents the *excess* number of SSD visits. This area roughly approximates the number of visits representing missed diagnostic opportunities. Therefore, some of the visits during the diagnostic opportunity window would also be expected to occur based on trends prior to this period (i.e., visits shaded in red to the right of the dashed-grey line), and are, thus, not considered missed opportunities. For example, we would expect some patients with tuberculosis to, coincidentally, have had fever within 90 days prior to their tuberculosis diagnosis.

Given estimates for the observed and expected trends during the diagnostic opportunity window, one can approximate the *number of missed opportunities* on a given day by subtracting the number of expected SSD visits from the number of observed SSD visits (either the true number or estimated trend in observed visits, as depicted by the blue line, may be used in this context.) The observed and expected trends depicted in Fig. 2 can be estimated in a variety of ways (e.g., linear or non-linear curves or non-parametric approaches); below we describe examples and considerations for estimating these trends. This and similar approaches using observed and expected visits have been used in prior investigations of diagnostic delays [17–19].

Bootstrapping Approach to Identify Likely Missed Opportunities

The above framework may be used to estimate the number of missed opportunities each day during the diagnostic opportunity window. However, because some visits represent *expected* SSD-related visits during the diagnostic opportunity window (i.e., the red shaded region in Fig. 2), it is not possible to directly identify which individual patient visits represent a missed opportunity from observational data alone. From the estimated number of missed opportunities alone we cannot determine: (1) the number of patients who experienced a missed opportunity, (2) the typical duration of diagnostic delays, nor (3) the number of missed opportunities that a typical patient experienced. Moreover, it may be challenging to estimate risk factors for experiencing a missed opportunity if individual visits representing a diagnostic opportunity cannot be identified. We refer to these types of measures as *patient-level metrics* associated with diagnostic delays. We propose a simulation framework designed to estimate each of these patient-level metrics

using a bootstrapping-based approach to resample the observation representing a likely missed opportunity. We do so by repeatedly sampling (i.e., randomly selecting) which visits represent a missed opportunity and then computing the individual-patient-level metrics of interest.

Let m_t denote the number of estimated missed opportunities at each day $t \in \{w, w+1, \ldots, -2, -1\}$ during the diagnostic opportunity window, where $w < 0$ denotes the point representing the start of the diagnostic opportunity window (see dashed grey line in Fig. 2).

Below we describe three different algorithms that may be used to simulate (i.e. sample) missed visits. In general, these approaches can be described by the following steps. Given estimates for m_t and w, described above, do the following:

1. For each time period in the diagnostic opportunity window, $t \in \{w, w+1, \ldots, -2, -1\}$, randomly select the number of missed visits m_t and label these as *missed opportunities.*
2. Aggregate all visits and corresponding patients who were drawn to represent a missed opportunity. Compute the number of patients missed, duration of delay (the time between first missed visit and the index date) for each patient, and the number of missed opportunities drawn for each patient.
3. Repeat steps 1 and 2 multiple times.
4. Aggregate results.

The following algorithms expand upon the selection procedure described in step 1 by preferentially drawing patient visits in relation to their perceived probability of representing a delay.

Algorithm 1: Independent draws

The first approach draws visits representing *missed opportunities* independent of one another at each time period in the diagnostic opportunity window. A formal description of this algorithm is presented in Fig. 3. This represents the simplest way to bootstrap missed opportunities but provides no correlation structure between the patients or visits that are selected at subsequent time points. For example, a patient who is drawn to have a missed opportunity because of symptoms occurring at 21 days before diagnosis would be no more likely to be drawn if they presented with symptoms 14 days prior to diagnosis.

Algorithm 2: Preferential selection of previously drawn cases

SSD visits occurring near the index diagnosis may be more likely to represent a missed opportunity if that patient also experienced a missed opportunity at earlier points before the index diagnosis. For example, if a patient has an SSD visit 15 days prior to the index diagnosis and then has another SSD visit at 5 days prior to the index, these visits are more likely to be missed opportunities than a single SSD visit.

Algorithm 1: A simple algorithm to draw potential missed opportunities at random across time periods

For a delay opportunity window between w and -1 periods prior to the initial diagnosis, where w is the start of the diagnostic opportunity window, we draw a set of missed visits V_t at each time period $t \in \{w, w+1, ..., -2, -1\}$ prior to the index diagnosis.

Let m_t denote the estimated number of missed visits to be drawn at each period t;

for *periods* $t \in \{w, w-1, ..., -2, -1\}$ **do**

  randomly select m_t of the SSD visits that occurred at period t, these visits become V_t;

end

identify the set of patients S who had a miss visit in the set $\{V_w, V_{w+1}, ..., V_{-1}\}$, for each patient in S compute the total number of missed visits and the duration of delay as the maximum time period in which they had a missed visit.

Fig. 3 Simple algorithm to simulated missed opportunities using uncorrelated draws

Thus, a second approach is to preferentially sample visits from patients who have been drawn at earlier time points in the diagnostic opportunity window. Figure 4 presents a formal description of an algorithm to select patient visits in a correlated fashion. This algorithm introduces a scaling parameter allowing one to regulate the preference given to selecting patients who have previously been drawn. Specifically, given a scaling parameter $\alpha \in [0,1]$, $m_t * (1-\alpha)$ of the visits at time point t will be selected from patients with previous missed visits (if available) and $m_t * \alpha$ of the visits at time point t will be selected from patients not previously drawn to have a missed opportunity. Note, a value of $\alpha = 0$ denotes strict preference to previously drawn patients while $\alpha = 0.5$ denotes equal preference, and $\alpha = 1$ denotes strict preference to patients not drawn at prior time steps. Using this selection procedure and given a set of estimates $\{m_t | t \in [w, -1]\}$ for missed visits,$\alpha = 0$ will minimize the number of patients with at least one missed opportunity, and sample patients with densely populated missed opportunities during the diagnostic opportunity window. In contrast, $\alpha = 1$ will maximize the number of patients with at least one missed opportunity, and sample patients with sparsely populated missed opportunities during the diagnostic opportunity window.

Algorithm 3: Generalized algorithm

We may also want to preferentially select visits from patients who are more likely to represent a missed opportunity based on multiple criteria. For example, a patient who has multiple healthcare encounters with unresolved symptoms may be more likely to represent a missed diagnostic opportunity. Similarly, a patient who presents with multiple different symptoms may be more likely to represent a diagnostic delay. Figure 5 presents an example of a more generalized algorithm that allows multiple criteria to be incorporated into the preferential selection criteria. This algorithm incorporates a functional weighting parameter based on the number of times a patient had SSD visits labeled as a missed opportunity and the number of distinct symptoms/SSDs the patient experienced during the current visit or prior visits in the SSD window.

R Package

We have developed an R package to implement the above algorithms. This package can be found at https://github.com/aarmiller/delaySim along with installation instructions and examples.

Considerations for Estimating Simulation Parameters

In this section, we discuss possible approaches to estimate the primary parameters (i.e., number of missed opportunities each day) necessary to implement the simulations described above. It is important to note that the bootstrapping approaches described above are indifferent to the estimation procedure used to estimate the number of missed opportunities. This section is intended to provide examples that guide the estimation process, but careful considerations of the methodological assumptions should be given to the specific clinical application (see Supplementary Methods S1 for further details). We begin by describing approaches to estimate the number of missed opportunities each day during the diagnostic opportunity window. We then describe approaches to estimate the bounds of the diagnostic opportunity window. Depending on the approach one chooses, these two parameters may be estimated simultaneously.

Estimating the expected trend and number of missed opportunities (m_t)

Let $t \in -T, -T+1, \ldots, -2, -1$ represent time points prior to the index diagnosis, where $-T$ represents the maximum amount of time prior to diagnosis that we wish to analyze. Our goal is then to estimate $y_t = f(t) + \epsilon_t$, where y_t is the number of SSD visits at time t over the interval $[-T, w-1]$. Then using this estimate we extrapolate $\widehat{y}_t$ to the interval $[w, -1]$.

Alternatively, we can specify the estimation problem over the entire interval as a piecewise function as follows:

$$y_t = \begin{cases} f(t) & \text{for } t < w \\ g(t) & \text{for } t \geq w \end{cases}$$

where $f(t)$ is the trend in SSD visits prior to the diagnostic opportunity window, and $g(t)$ is the trend in SSD visits during the diagnostic opportunity window. A variety of model fitting approaches may be used to estimate $f(t)$ and/or $g(t)$. For example, Fig. 6 depicts a case where $f(t)$ is either a linear function of time (left) or an exponential function (right). Similarly, various time-series modeling approaches may be used to capture temporal aspects of the estimation problem (e.g., periodicity, autocorrelation).

In general, we have found in previous work [18, 19, 24, 25] that the trend prior to the diagnostic opportunity window ($f(t)$) can be roughly approximated by a linear model, while the trend during the diagnostic opportunity window ($g(t)$) is typically non-linear. It is worth emphasizing that the choice of functional form for $f(t)$ will impact the number of visits during the diagnostic opportunity window considered to be a missed opportunity, as depicted in Fig. 6. A linear specification may lead to fewer visits falling below the expected trend while a non-linear specification may lead to a greater proportion of visits falling below the expected trend (see Supplementary Methods S1 for further details).

Once values for $\widehat{f}(t)$ have been obtained, we can compute the number of missed opportunities at a given time

Algorithm 2: An algorithm to preferentially select potential missed opportunities from patients previously labeled as being missed

For a delay opportunity window between w and -1 periods prior to the initial diagnosis, we draw a set of missed visits V_t at each period t starting at the period w and working in towards period -1. We identify the corresponding set of patients S_t where $\Phi : V \mapsto S$ maps a set of visits on a given date to the corresponding set of patients.

Let m_t denote the estimated number of missed visits to be drawn at each period $t \in \{w, w-1, ..., -2, -1\}$;

Let $\alpha \in [0, 1]$ denote a scaling parameter that determines the preference given to selecting visits from patients that previously had visits selected as being missed, where $\alpha = 0$ denotes strict preference is given to previously drawn agents (i.e., previously drawn agents are selected prior to new agents) and $\alpha = 1$ denotes strict preference to previously undrawn agents;

for *periods* $t \in \{w, w+1, ..., -2, -1\}$ **do**

  identify the set of visits Ω_t containing an SSD diagnosis in period t, and the set of patients $\Theta_t = \bigcup_{j=w}^{t-1} S_j$ previously drawn as having a missed visit up to period t. Divide Ω_t into two subsets, among patients with a prior missed visit $\Omega_t^1 = \{x | x \in \Omega_t, \Phi(x) \in \Theta_t\}$ and those without a prior missed visit $\Omega_t^2 = \{x | x \in \Omega_t, \Phi(x) \notin \Theta_t\}$;

  if $|\Omega_t^1| \geq m_t(1-\alpha)$ *and* $|\Omega_t^2| \geq m_t\alpha$ **then**

    randomly select $m_t(1-\alpha)$ visits from Ω_t^1 and $m_t\alpha$ visits from Ω_t^2, these visits become V_t and the corresponding set of patients $\Phi(V_t)$ become S_t;

  else if $|\Omega_t^1| < m_t(1-\alpha)$ **then**

    select all visits from Ω_t^1 and randomly select $m_t - |\Omega_t^1|$ visits from Ω_t^2, these visits become V_t and the corresponding set of patients $\Phi(V_t)$ become S_t;

  else

    select all visits from Ω_t^2 and randomly select $m_t - |\Omega_t^2|$ visits from Ω_t^1, these visits become V_t and the corresponding set of patients $\Phi(V_t)$ become S_t;

end

analyze the final set of SSD visits $\{V_w, V_{w+1}, ..., V_{-1}\}$ associated with the set of missed patients $\{S_w, S_{w+1}, ..., S_{-1}\}$ at each corresponding time point, respectively, for each patient in $S = \bigcup_{j=w}^{-1} S_j$ compute the total number of missed visits and the duration of delay as the maximum time period in which they had a missed visit.

Fig. 4 An algorithm to draw patients with preference given to patients previously drawn

Algorithm 3: A generalized algorithm to preferentially select potential missed opportunities from patients with SSD patterns more likely to represent a delay and previously labeled as being missed

For a delay opportunity window between w and -1 periods prior to the initial diagnosis, we draw a set of missed visits V_t at each period t starting at the period w and working in towards period -1.

Let $\Phi : V \mapsto S$ map a set of visits to the corresponding set of patients;

Let m_t denote the estimated number of missed visits to be drawn at each period $t \in \{w, w+1, ..., -2, -1\}$;

Let $SSD_{i,t}$ be the set of distinct SSDs that patient i experienced at time t (or between $[w, t]$), let $\mathbb{C}_{i,t}$ represent a count of the number of times individual i has been selected for having a missed opportunity in period $[w, t)$, and let $w_t(SSD_{i,t}, \mathbb{C}_{i,t-1})$ be a function that assigns a sampling weight to the set of patient visits at time t as a function of $SSD_{i,t}$ and $\mathbb{C}_{i,t-1}$

for *periods* $t \in \{w, w-1, ..., 0\}$ **do**

identify the set of visits Ω_t containing an SSD diagnosis in period t;

compute $SSD_{i,t}$ and $\mathbb{C}_{i,t-1}$ for each $i \in \Phi(\Omega_t)$;

randomly select m_t visits from from Ω_t where each visit is selected with the probability $w_t(SSD_{i,t}, \mathbb{C}_{i,t-1}) / \sum_{j \in \Omega_t} w_t(SSD_{j,t}, \mathbb{C}_{j,t-1})$;

the set of selected visits are assigned to V_t and the corresponding set of patients $S_t = \Phi(V_t)$.

end

analyze the final set of SSD visits $\{V_{-w}, V_{-w+1}, ..., V_{-1}\}$ associated with the set of missed patients $\{S_w, S_{w+1}, ..., S_{-1}\}$ at each corresponding time point, respectively, for each patient in $S = \bigcup_{j=-w}^{-1} S_j$ compute the total number of missed visits and the duration of delay as the maximum time period in which they had a missed visit.

Fig. 5 A generalized algorithm to draw patients with preference to previously drawn patients and those with multiple symptoms

point t in one of two ways. First, if an explicit value for $\widehat{g}(t)$ has not been obtained, we can use the observed count at time t, such that $m_t = y_t - \widehat{f}(t)$. Second, we can use the fitted value for $\widehat{g}(t)$ such that $m_t = \widehat{g}(t) - \widehat{f}(t)$. Confidence bounds around the number of missed opportunities may also be computed using appropriate prediction intervals around $\widehat{f}(t)$ or $\widehat{g}(t) - \widehat{f}(t)$.

Estimating the bounds of the diagnostic opportunity window (change-point detection)

The lower bound of the diagnostic opportunity window, w, represents the cross-over point prior to diagnosis used to delineate the diagnostic opportunity window from the period where the expected pattern of care is estimated. Thus, w must be defined before calculating m_t as noted above. While this bound on the diagnostic opportunity window may be specified *a priori* based on clinical knowledge, it may also be desirable to estimate this "change point" as part of the analytical process.

One approach for finding this change point is to employ standard change-point finding algorithms to find the optimal value for $cp = w$ using the trends outlined above. For example, given a parametric specification for $f(t)$ and $g(t)$, one approach to find the optimal value

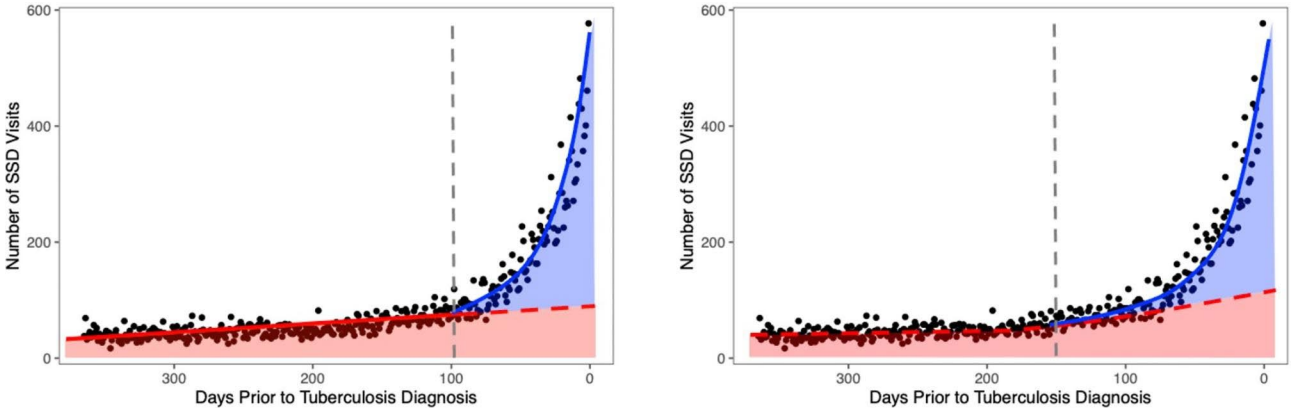

Fig. 6 Estimating expected number of visits using linear (left) or exponential (right) curves to represent the expected number of SSD visits

cp may be achieved by iterating over different values for cp and comparing the fitted performance of $\widehat{f}(t)$ and $\widehat{g}(t)$ (e.g., by minimizing the Akaike or Bayesian information criterion [AIC/BIC] or minimizing mean squared error [MSE], etc.) Change-point-detection approaches may also be used that do not require explicit specification of functional forms for $f(t)$ or $g(t)$ such as the CUSUM method.[27] Alternatively, if using the formula $m_t = y_t - \widehat{f}(t)$ to estimate the number of missed opportunities without estimating $g(t)$, it is possible to find cp by exploiting the assumption that $y_t > f(t), \forall t > cp$. In Supplementary Methods 2, we present one such approach that we later refer to as the *prediction bound* approach.

Applications

We apply our bootstrapping approach to three diseases where similar retrospective approaches have been previously applied to study diagnostic delays: stroke, AMI, and tuberculosis (TB). We chose these diseases because they generally present with different symptoms, require different diagnostic testing approaches, and the possible diagnostic delays associated with these diseases have different durations. Strokes primarily present with neurologic symptoms (e.g., weakness);[9, 12, 13] AMIs present with cardiac symptoms (e.g., chest pain and pressure);[8, 10, 11, 17, 28–31] and TB presents with respiratory symptoms (e.g., cough) [16, 18]. For each of these conditions, we use criteria from these prior studies to identify case patients and the index diagnosis. We then compute the number of SSD-related visits each day prior to the diagnosis and use these counts to estimate the number of missed opportunities. Supplementary Table 2 describes the diagnosis codes and sources used to identify each index condition. Supplementary Table 2 describes SSDs used for each condition and their corresponding diagnosis codes (note, only ICD-9 and ICD-10 diagnosis codes were used to identify SSDs). Each of these criteria were selected based on prior investigations of diagnostic delays for these diseases [8, 9, 13, 16–18, 28–31].

Study Population. We used administrative claims data from the IBM MarketScan Commercial Claims Databases from 2001 to 2017. This database contains longitudinal insurance claims for individuals with employer-sponsored health insurance along with spouses, partners and dependents of the primary enrollee. Over this study period, records are available for over 185 million distinct enrollees. Claims for inpatient, outpatient, emergency department (ED) and prescription medications are provided. These include diagnosis and procedure codes, outpatient medications, admission and discharge characteristics, and enrollment and demographic information.

For each of the study conditions, we identify the first time an enrollee was diagnosed with the disease of interest and label this as the index diagnosis. We exclude children < 18 years of age and enrollees with less than 365 days of continuous enrollment prior to the index diagnosis. Based on prior investigations and clinical plausibility, delays associated with tuberculosis can be expected to exceed a few months while delays for AMI and stroke are typically a month or less. Thus, we used a value of $T = 365$ days for tuberculosis and $T = 180$ days for stroke and AMI.

Estimation procedures for simulation parameters. For each condition, we compare three approaches to find the potential change-point. First, we fit a piecewise linear-cubic model with a linear trend over the interval $[-T, cp - 1]$ and a cubic trend over the period $[cp, -1]$. We iterate over values for cp and choose the optimal value based on AIC. Second, we used the CUSUM method to detect the change point over the interval prior to the index diagnosis beginning at $-T$. For this approach we use a linear model to estimate the expected trend during visits prior to the identified change-point cp. Third, we use the prediction bound approach, described in Supplementary Methods 2, to identify the point where the observed values are systematically greater than the

95% prediction bound during the diagnostic opportunity window. For this approach, we also use a linear model to estimate the trend in expected SSD visits. Finally, we select the optimal change-point method based on the consistency of the corresponding model fit with the data generating process outlined above (see Supplementary Methods 3).

After selecting the optimal change-point approach, we then estimate the number of missed opportunities each day during the window. Specifically, we estimate the expected trend using a simple linear function of time prior to diagnosis t by fitting the model $y_t = \beta_0 + t\beta_1 + \epsilon_t$ over the interval $[-T, cp-1]$ and extrapolating $\widehat{y}_t$ over the interval $[cp, -1]$. We then compute the number of missed visits as the observed error $m_t = y_t - \widehat{y}_t$ using the observed count y_t over the interval $[cp, -1]$.

Bootstrapping estimates. We compare three different simulation approaches using two of the algorithms described above. First, we use Algorithm 1 to draw missed opportunities that are independent at each period. Second, we use Algorithm 2 while setting $\alpha = 0$. Third, we use Algorithm 2 while setting $\alpha = 1$. Thus, our first simulation approach will describe the expected number of patients with a delay using uncorrelated draws, while the second and third models will roughly summarize the minimum and maximum number of patients potentially experiencing a delay, respectively, using correlated draws. For simplicity, we refer to these approaches as *naïve* draws (algorithm 1), *dense* patient-delays (algorithm 2, $\alpha = 0$) and *sparse* patient-delays (algorithm 2, $\alpha = 1$). We use the terms dense and sparse to denote that $\alpha = 0$ will tend to draw patients with densely populated opportunities during the diagnostic opportunity window, versus patients with more sparsely populated opportunities for $\alpha = 1$.

Using these three simulation approaches, we estimate the following measures of the frequency of missed opportunities: the total number of visits representing a missed opportunity; the percent of missed opportunities occurring in inpatient, outpatient and emergency department settings; the percent of patients experiencing a missed opportunity; the mean number of missed opportunities each patient experienced; the mean duration of diagnostic delays (time from earliest missed opportunity to index diagnosis). We compute bootstrap-based 95% confidence intervals for each of these estimates by repeatedly redrawing 1,000 times which visits represented a missed opportunity, computing the above metrics, and using the 2.5 and 97.5 percentile values.

Sensitivity analyses. It is possible that our SSD list and corresponding SSD visits may not fully capture all visits where a patient presented with symptoms of the disease. For example, there may be other SSD codes that are not directly specified in our SSD list. Alternatively, symptoms that occur during a visit may not be captured in the administrative discharge record because of recording errors (e.g., clinician fails to record symptom) or due to billing issues (e.g., no corresponding ICD code applied to the insurance claim). To address this potential limitation, we repeat our change-point analysis, and employ our simulation models for all visits, instead of SSD visits only. This provides an upper bound on the number of potential missed opportunities.

As a second sensitivity analysis, we evaluate how the selection of a non-linear expected SSD curve may influence the estimated number of missed opportunities. Specifically, for each of the selected change points, we also consider a cubic trend to approximate this baseline period of expected SSD visits. We then compute the difference in the expected number of missed opportunities that may be used to inform the bootstrapping approach.

Results

We identified 2,073 cases of tuberculosis, 359,625 cases of AMI, and 367,768 cases of stroke. Table 1 presents baseline characteristics for each of our study populations, including demographics, enrollment information and the number of observable visits per patient during the observation period prior to the index diagnosis. Figure 7 presents counts of SSD visits for each day leading up to the index diagnosis for each condition. There is a significant increase in SSD-related visits for all three conditions that occurs in the period before diagnosis.

For each condition, we estimated the change point, expected number of baseline visits, and the number of missed visits using 3 different change-point detection approaches. Supplementary Tables 3 and Supplementary Figs. 1–3 present the resulting change points, summaries of the number of missed opportunities and a visualization of the diagnostic opportunity window and expected number of SSD visits. For each condition, we then identified the change-point approach that appeared to best fit the pattern of SSD visits; Supplementary Table 4 lists the evaluation criteria that were used. Note: each of these change points that were identified are consistent with previous research.

For tuberculosis we selected the linear-cubic model which identified the change-point at 114 days prior to diagnosis. For AMI we selected the prediction bound model, which resulted in a change-point at 40 days before diagnosis. For stroke we also selected the prediction bound model, which resulted in a change-point at 39 days before diagnosis. Table 2 summarizes these results for each condition and the final selected change-point approach. Figure 7 also provides visualizations of the SSD visit counts along with the expected trend and corresponding opportunity window. Supplementary Table 5 provides the final estimates of the number of missed

Table 1 Baseline characteristics of the study cohorts for Stroke, AMI and Tuberculosis used for evaluating our simulation approach

	Stroke	AMI	Tuberculosis
N	367,768	359,625	2,073
Age at Diagnosis (n (%))			
18–30	17,972 (4.89%)	4,174 (1.16%)	252 (12.16%)
31–45	61,820 (16.81%)	45,835 (12.75%)	620 (29.91%)
46–55	121,995 (33.17%)	131,604 (36.59%)	588 (28.36%)
>55	165,981 (45.13%)	178,012 (49.50%)	613 (29.57%)
Sex (n (%))			
Male	186,966 (50.84%)	249,899 (69.49%)	988 (47.66%)
Female	180,802 (49.16%)	109,726 (30.51%)	1,085 (52.34%)
Enrollment Time Prior to Index (years)			
Mean	3.62	3.70	3.86
Median	2.60	2.65	2.87
Range	0.49–17.01	0.49–17.01	1.00–15.79
Count ≥ 1 year (n (%))	306,800 (83.42%)	302,026 (83.98%)	2,073 (100.00%)
Count ≥ 1.5 years (n (%))	260,581 (70.85%)	257,661 (71.65%)	1,719 (82.92%)
Count ≥ 2 years (n (%))	221,360 (60.19%)	219,651 (61.08%)	1,419 (68.45%)
Count ≥ 3 years (n (%))	162,047 (44.06%)	161,224 (44.83%)	997 (48.09%)
Number of Visits in Period Prior to Diagnosis (i.e. 365 days prior for tuberculosis and 180 days prior for stroke and AMI) (n (%))			
0	52,737 (14.34%)	63,141 (17.56%)	52 (2.51%)
1–5	124,630 (33.89%)	137,677 (38.28%)	251 (12.11%)
6–10	71,075 (19.33%)	66,292 (18.43%)	345 (16.64%)
11–20	64,189 (17.45%)	52,606 (14.63%)	606 (29.23%)
> 20	55,137 (14.99%)	39,909 (11.10%)	819 (39.51%)

opportunities at each time period that were used in the simulation models.

For each condition we applied three different bootstrapping approaches (naïve, sparse, and dense), as described above, to sample which visits represented a missed opportunity. Table 2 presents the results of these simulations in terms of the settings where missed opportunities occurred (outpatient, inpatient, or ED), the percentage of patients experiencing at least one missed opportunity, the mean number of missed opportunities such patients experienced and the average duration of diagnostic delays. Supplementary Tables 6–8 provide expanded results for stroke, AMI, and TB, respectively.

In general, across all three diseases most diagnostic opportunities appeared to occur in outpatient settings, representing around 90% of missed opportunities for tuberculosis, 75% for AMI, and 60% for stroke. The ED was the second most common setting for misses representing around 30% for stroke, 20% for AMI, and 5% for tuberculosis. For all three diseases, around 4–6% of missed opportunities occurred in inpatient settings. The type of algorithm did not dramatically alter these estimates, with the percentage only differing by a percentage point or two across algorithms for each disease. However, the sparse approach did tend to result in fewer missed opportunities ascribed to outpatient settings, while the dense approach resulted in slightly more, and the naïve approach tended to be in between algorithms 2 and 3.

The differences between the algorithms are reflected more clearly in the estimates of the percentage of patients experiencing a missed opportunity, the number of missed opportunities ascribed to each patient who was missed, and the duration of diagnostic delays among patients who were delayed. The sparse approach tended to result in the largest number of patients experiencing a miss, as expected; the dense approach resulted in the fewest while the naïve approach was somewhere in between. Patients with tuberculosis were most likely to experience a missed opportunity, with between 63.9% (95% CI: 62.81–65.03) and 82.3% (CI: 81.96–82.59) of patients experiencing at least one missed opportunity, depending on the algorithm, while patients with stroke were least likely, with 6.9% (CI: 6.91–6.96) to 8.3% (CI: 8.28–8.31) of patients experiencing a missed opportunity across the different algorithms.

Similarly, the correlated nature of successive draws results in differences in the average number of missed opportunities assigned to patients who experience a delayed diagnosis. The dense approach identified the fewest number of patients with a missed opportunity and resulted in more missed opportunities assigned to each patient compared to the sparse approach, which tended to identify the greatest number of individuals with at least one missed opportunity. Patients with tuberculosis experienced the greatest number of missed opportunities, with between 3.78 (CI: 3.76–3.79) missed opportunities per patient for the sparse approach versus 4.87 (CI: 4.78–4.95) missed opportunities for the dense approach. Patients with AMI and stroke who experienced a missed opportunity had a similar number of missed opportunities across algorithms (1.36 [CI: 1.36–1.37] to 1.63 [CI:

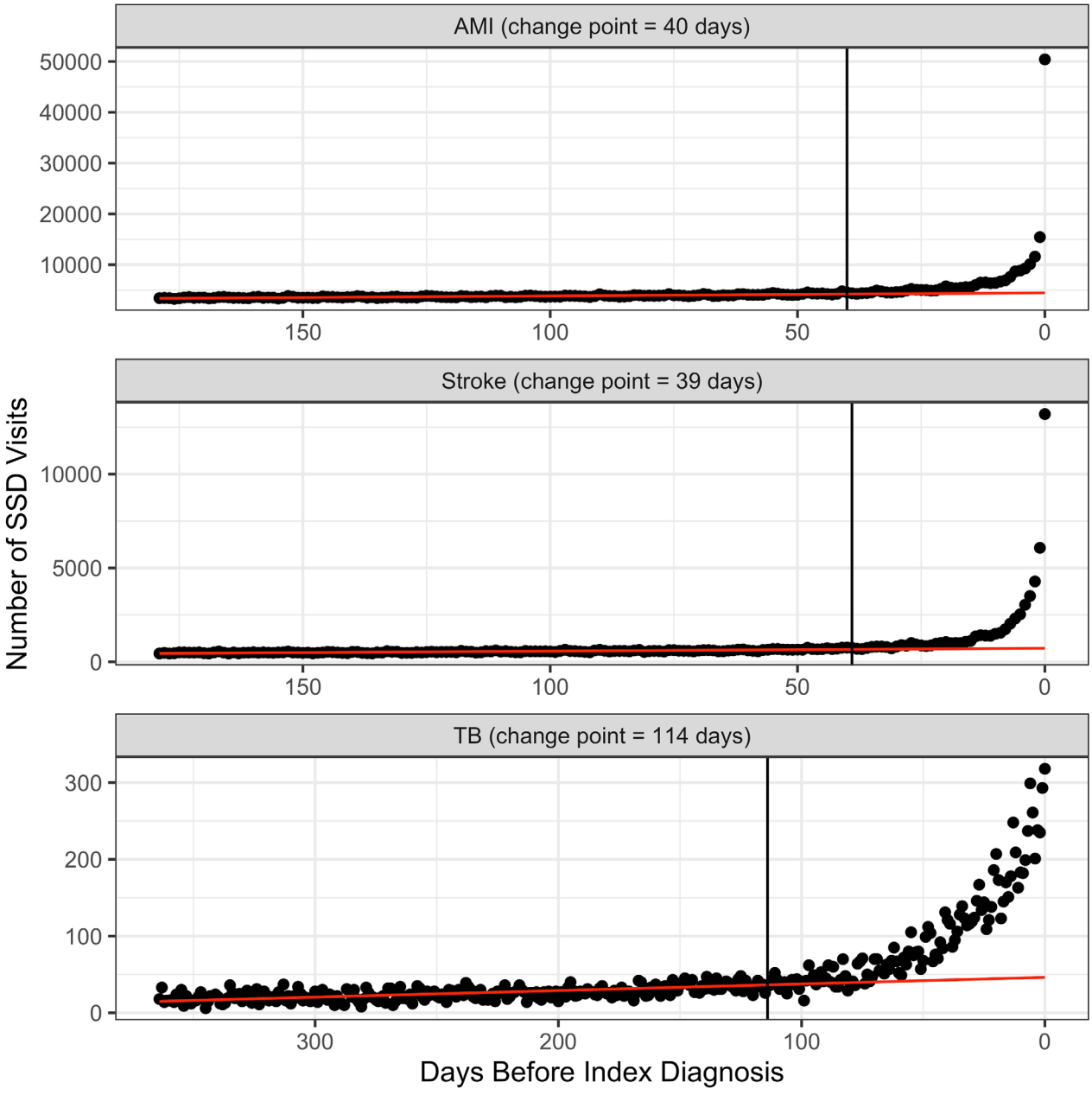

Fig. 7 Counts of SSD visits each day prior to the index diagnosis. For each disease of interest there is an upward spike in the occurrence of healthcare visits with SSDs in the period just preceding the index diagnosis. The black vertical line represents the estimated change-point separating the diagnostic opportunity window from the prior crossover period. The red line represents the expected level of healthcare utilization (i.e., estimated to occur in absence of diagnostic delays)

1.62–1.64] per patient for stroke and 1.36 [1.35–1.36] to 1.81 [CI: 1.81–1.82] for AMI).

A somewhat counterintuitive result occurs across algorithms related to the estimated duration of diagnostic delays. The dense approach tended to result in a shorter average duration of diagnostic delays, compared to the sparse approach which produced longer diagnostic delays on average. This result is due to the skewed nature of SSD visits (see Fig. 7) and the correlated nature of subsequent draws between the algorithms. Because the dense approach attempts to draw missed opportunities at each subsequent time period from patients who have already been drawn (i.e., a longer delay), there is a clustering of the earlier SSD visits among patients with already long duration of delays; those visits which occur further before the index date are more likely to be assigned to patients already having a longer delay, while newly drawn patients are more likely to be selected corresponding to

Table 2 Selected Simulation Results – Estimates of the frequency of missed opportunities and duration of delays using different simulation algorithms

	Stroke	AMI	Tuberculosis
Change Point (Start of diagnostic opportunity Window	39	40	114
Total Number of Missed Opportunities During Delay Window (% of SSD visits during delay window)	41,577 (60.55%)	117,344 (46.24%)	6,444 (58.15%)
Percent of Patients Experiencing a Delay of at least 1 day (95% CI)			
Naïve Approach	7.46 (7.43–7.48)	20.86 (20.81–20.91)	78.11 (77.33–78.87)
Dense Delays Approach	6.93 (6.91–6.96)	18.01 (17.96–18.07)	63.90 (62.81–65.03)
Sparse Delays Approach	8.29 (8.28–8.31)	24.08 (24.04–24.11)	82.28 (81.96–82.59)
Mean Number of Missed Opportunities among Patients Missed (95% CI)			
Naïve Approach	1.52 (1.51–1.52)	1.56 (1.56–1.57)	3.98 (3.94–4.02)
Dense Delays Approach	1.63 (1.62–1.64)	1.81 (1.81–1.82)	4.87 (4.78–4.95)
Sparse Delays Approach	1.36 (1.36–1.37)	1.36 (1.35–1.36)	3.78 (3.76–3.79)
Mean Duration (days) of Missed Opportunities among Patients Missed (95% CI)			
Naïve Approach	7.41 (7.38–7.44) Days	8.10 (8.08–8.12) Days	39.86 (39.31–40.37) Days
Dense Delays Approach	6.72 (6.67–6.76) Days	6.69 (6.64–6.73) Days	34.35 (33.52–35.15) Days
Sparse Delays Approach	7.64 (7.63–7.66) Days	8.21 (8.20–8.22) Days	44.47 (44.16–44.75) Days
Percent of Missed Opportunities in Outpatient Settings			
Naïve Approach	62.64 (62.42–62.88)	75.99 (75.84–76.14)	89.23 (88.83–89.66)
Dense Delays Approach	63.04 (62.80–63.30)	76.55 (76.39–76.72)	90.25 (89.82–90.72)
Sparse Delays Approach	61.52 (61.34–61.71)	74.76 (74.64–74.89)	88.58 (88.21–88.97)
Percent of Missed Opportunities in Inpatient Settings			
Naïve Approach	4.37 (4.25–4.48)	5.67 (5.58–5.76)	5.03 (4.72–5.31)
Dense Delays Approach	4.26 (4.15–4.37)	5.64 (5.56–5.73)	4.89 (4.61–5.17)
Sparse Delays Approach	4.50 (4.40–4.61)	5.68 (5.60–5.76)	5.07 (4.81–5.32)
Percent of Missed Opportunities in ED Settings			
Naïve Approach	33.00 (32.78–33.21)	18.34 (18.21–18.46)	5.74 (5.40–6.05)
Dense Delays Approach	32.70 (32.47–32.92)	17.81 (17.67–17.94)	4.86 (4.52–5.18)
Sparse Delays Approach	33.98 (33.80–34.16)	19.56 (19.45–19.6)	6.35 (6.07–6.60)

visits closer to the index date. Thus, the dense approach generates a relatively small number of patients who have long delays but with many visits further before the index date while producing relatively more patients with short delays closer to the index. This difference between algorithms has a relatively large impact on the average duration for diagnostic delays in the case of tuberculosis, with the average duration of 34.35 (CI: 33.52–35.15) days for the dense approach versus 44.47 (CI: 44.16–44.75) days for the sparse approach. However, the difference between algorithms is less dramatic for stroke (6.72 [CI: 6.67–6.76] vs. 7.64 [CI: 7.63–7.66]) and AMI (6.64 [CI: 6.60–6.68] vs. 8.16 [CI: 8.15–8.17]).

We also conducted a sensitivity analysis of the SSD list used to identify potential missed opportunities. We repeated all analyses using all visits, instead of SSD visits only. These results are presented in Supplementary Tables 9–10. In general, this analysis resulted in a much greater estimated number of missed opportunities (more than 5 times as many for stroke, 1.45 times for AMI and 1.52 times for tuberculosis). The general trends across algorithms, described above, were the same with three notable exceptions. First, a greater proportion of missed opportunities were estimated to occur in outpatient settings. Second, a greater percentage of patients were estimated to have experienced a missed opportunity; this result was most exaggerated for AMI and stroke. Third, the mean number of missed opportunities and duration of delays among those patients who experienced a delay was not consistently different; in some cases, the number of delays per patient and duration of delays increased, while in others it was the same or even decreased. Thus, this type of sensitivity analysis may help to broaden the number of missed opportunities and patients with missed opportunities identified but may not dramatically change the dynamics of missed opportunities among patients identified to have a delay in terms of the number of missed opportunities they are estimated to experience or the estimated duration of delay.

As a second sensitivity analysis, we compared how the baseline trends in expected SSD visits would be altered if the same change points were used but non-linear (i.e.,

cubic) baseline curves were used to estimate the expected trend in SSD visits. Supplementary Table 11 provides the results. In general, the use of a non-linear expected trend resulted in slightly fewer estimated missed opportunities during the diagnostic opportunity window. Thus, if there is reason to suspect this unobserved period of activity is non-linear, results estimated using a linear trend may tend to overestimate the number of missed opportunities.

Discussion

In this paper, we presented a general bootstrapping-based simulation approach to estimate individual-level measures of missed diagnostic opportunities by resampling observed healthcare visits from longitudinal health records. Specifically, this approach allows one to estimate the frequency of missed opportunities at an aggregate level, along with individual-level metrics such as the number of patients experiencing a missed opportunity, the settings where missed opportunities occur, the number of missed opportunities that individual patients experience, and the duration of diagnostic delays. We applied these methods to TB, AMI, and stroke, and consistent with prior investigations for these diseases, we identified a significant number of missed opportunities associated with these diseases. We also demonstrated that a range of results may be generated based on different evaluation criteria.

The bootstrapping approach we describe, unlike many prior approaches to study diagnostic delays, such as retrospective chart reviews, autopsy studies, or malpractice claims is less costly and time consuming and can be applied to virtually any disease captured by longitudinal patient records. Moreover, our approach provides a high degree of flexibility in terms of estimation procedures, sampling algorithms for selecting missed opportunities, and output measures. While other approaches have used similar longitudinal data sources and a type of retrospective approach to study diagnostic delays, these approaches often have methodological limitations that our approach is designed to address [8, 9, 11, 13, 16]. First, these studies are often unable to compute individual-level patient metrics, such as delay duration or frequency of misses in individual patients, as these metrics typically require analyzing individual patient records. Second, these studies generally require expert specification of criteria to define a delay (e.g., time prior to diagnosis). Third, the approaches very often do not account for the fact that many signs and symptoms occurring before diagnosis may be unrelated to the disease of interest and can be expected to occur even in absence of diagnostic delays; thus, prior approaches may not distinguish between likely missed opportunities and coincidental visits. To our knowledge, only a few prior studies have attempted to control for this coincidental level of care[17–19] and our approach can be viewed as an extension of these prior methods.

We presented three sampling algorithms and applied two of these to the diseases of interest. Given our findings, which demonstrate differing results across algorithms, we offer the following guidance to future investigators wishing to use this approach. First, in cases where relatively little is known about the correlation between patient revisits and the likelihood of an individual visit representing a delay, we recommend the naïve approach (Algorithm 1) as the baseline or default estimate. This algorithm places the fewest assumptions on the simulation process and is conceptually the most straightforward. However, investigators may also consider Algorithm 2, while setting $\alpha = 0$ and $\alpha = 1$ to provide bounds on the results. Second, in situations where more information is known about the correlation between repeated missed opportunities prior to diagnosis (e.g., symptoms are known to persist among patients who are delayed), expert evaluation may be used to determine if α, in Algorithm 2, should be set closer to 0 or 1. Finally, as future investigations make use of these approaches, and validation studies are conducted, such information may be used to develop a more realistic specification for the generalized algorithm (Algorithm 3) customized to specific diseases.

One of the most important contributions of our approach is that we attempt to account for coincidental healthcare that may occur prior to the index diagnosis of a disease of interest, while also providing a means for generating individual-level delay analysis. For example, not all respiratory events that occur prior to a tuberculosis diagnosis may be a direct result of tuberculosis, and many visits that appear to be potential missed opportunities may be coincidental. Failure to account for these expected trends may lead to significant overestimates in the frequency of diagnostic opportunities. However, prior investigations that have used similar data sources and approaches to identify individual missed diagnostic opportunities have typically labeled all events that meet pre-specified criteria (e.g., dizziness before stroke) as a "missed opportunity." Consequently, these approaches are often paired with additional criteria (e.g., treat and release ED visits [8, 9, 11]) to ensure greater specificity but come at the cost of decreasing the sensitivity in identifying missed opportunities. Attempts have been made to account for observed patterns of symptomatic visits relative to what would have been expected using either other visits [11] or using a crossover period prior to when delays may be expected to occur,[17] as we have proposed. However, these approaches do not yield individual-level-patient metrics, such as delay duration or frequency of misses in individual patients.

Another primary advantage of our approach is that it provides a flexible set of criteria for guiding the estimation process of diagnostic delays. First, the general bootstrapping approach we propose is agnostic to the methodological technique used to estimate the aggregate number of missed opportunities. Numerous procedures can be used to estimate the simulation parameters described above, including the change-point for the diagnostic opportunity window and the trend in expected SSD visits. Indeed, we compared three different change-point methods along with both linear and non-linear expected trends. Second, our generalized algorithm presents a customizable weighting parameter that can be used to adapt the simulation to a particular disease. Other extensions are possible, and the algorithms here can be customized for more complex scenarios. For example, a sequential selection criterion such as – *a patient who experiences SSD A then SSD B is more likely to represent a diagnostic delay than SSD B before SSD A*. Thus, our results present a foundation upon which future investigations can build.

Results from our different simulation approaches demonstrate that a range of estimates may be generated based on how one chooses to define the correlation structure between missed opportunities identified across sequential draws within the simulation. For example, the type of algorithm selected resulted in differences in the percent of individuals identified to have a delay, the setting where missed opportunities occurred (inpatient, outpatient, ED), the duration of delay and number of missed opportunities per patient. In some cases (e.g., mean duration of delays with tuberculosis), the difference in estimates between algorithms can be quite large. In addition, the selection of change-point approach and methods used to estimate the trend in expected SSD visits can impact the estimated number of missed opportunities used to conduct the bootstrapping analysis. Thus, researchers must be cognizant of when metrics generated may tend to reflect over- or under-estimates, as described in Supplementary Methods 1. Clinical knowledge must still be employed to guide both the estimation process and select the simulation technique that best suits the particular disease. However, the bootstrapping approach we present can provide bounds on the range of plausible results at an individual level. In addition, our sensitivity analysis using all visits prior to diagnosis may be one approach to provide an upper bound on the estimated number of missed opportunities.

A final benefit of our bootstrapping approach is that it may provide a unified framework for quantifying and comparing the frequency and duration of diagnostic delays across a variety of diseases in a more reproducible fashion. As noted above, there are considerable challenges when attempting to compare estimates of diagnostic delays across studies where differing methods and study populations are used. Our approach may provide a recourse for generating comparable results across diseases and studies, allowing investigators to directly compare which diseases may have a longer or shorter delay process and/or a greater/lesser frequency of delays. For example, in our applications, missed opportunities were far more common for tuberculosis compared to stroke and even less common for AMI. Average delays for AMI and stroke were similar at around 7–8 days versus tuberculosis, which was around 40 days. For all three diseases, only around 5% of the missed opportunities we identified occurred in inpatient settings, but there were considerable differences between ED and other outpatient care. These and similar metrics may also be useful for benchmarking purposes or providing a measure of *diagnostic efficiency* (e.g., how many healthcare resources are typically required to make a correct diagnosis) across diseases. Such measures may be useful for policy makers wishing to evaluate the relative importance of delays across a wide range of diseases using widely available data sources.

Our bootstrapping approach represents a technique for estimating the potential frequency and duration of diagnostic delays for a given disease. However, these methods do not directly measure nor evaluate the outcomes of a given delay in terms of morbidity or mortality. The harms associated with a diagnostic delay are likely to vary considerably across different diseases. For some diseases (e.g., TB) a delay of only a few days may not significantly impact outcomes, but for other diseases (e.g., stroke), delays of a few days could be catastrophic. Thus, additional clinical considerations are necessary when using measures of diagnostic delay to evaluate diagnostic error. Moreover, researchers may wish to manually adjust the criteria used in the algorithms described here to account for the clinical significance for a given delay [16, 18].

There are some limitations with the simulation approach we present here. First, our approach generally requires a large data source of longitudinal patient records, especially for diseases that are relatively rare. While such records are often readily available in the form of administrative claims, discharge records or other observational data sources, such data typically do not contain the types of granular information necessary for in-depth validation of delays (e.g., clinic notes or vital signs). A second limitation of our approach is that the missed opportunities identified do not necessarily imply diagnostic delays have occurred, and even "likely" missed opportunities for diagnosis may be unavoidable even in settings of ideal patient care. Our approach is simply designed to detect missed opportunities based on "excess" SSD visits that deviate from expectations to a statistically meaningful degree. However, we cannot assume

that a healthcare provider would reasonably be expected to diagnose each of these cases, and our approach does not incorporate harms that may have resulted. Similarly, our approach may miss longer delays, those that do not generate a significant signal in aggregate visit counts, or those that occur before the detected opportunity window. For example, the upward trend before diagnostic opportunity window may be partially driven by longer-duration delays. Thus, there may exist some patients whose delay is not completely captured by our approach. Finally, the approach outlined above assumes the defined SSD set is relatively complete. If a significant number of SSDs are unknown or not included in the primary analysis, results may be significantly underestimated. However, we also presented a sensitivity analysis using all visits that may be used to compute upper bounds on the number of delays and provide guidance on the relative completeness of the defined SSD set.

Conclusions

The bootstrapping-based simulation approach presented here provides an intuitive, flexible, and broadly applicable framework that can be used to identify missed opportunities and study diagnostic delays using large longitudinal data sources. This approach is less costly and time intensive than traditional methods to study diagnostic delays. It builds upon recent efforts to utilize large real-world datasets to study diagnostic delays, but also addresses many of the limitations present in prior study designs. Our results demonstrate consistency with prior investigations of diagnostic delays, but also provide a means to generate future results for different diseases and study populations. Moreover, we outlined several extensions upon which future investigations and clinical expertise may be used to expand and refine our general approach to individual diseases.

Abbreviations

SSD	Symptomatically Similar Diagnosis a diagnosis or disease that is a symptom of or shares symptoms to the disease of interest
CPT	Current Procedural Terminology
ICD	International Classification of Disease
TB	Tuberculosis
AMI	Acute Myocardial Infarction
AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
MSE	Mean Squared Error
CUSUM	Cumulative Sum Control Chart a sequential analysis technique that can be used to detect a statistical change-point

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12911-023-02148-w.

Supplementary Material 1

Acknowledgements

Not Applicable.

Author Contribution

ACM contributed to the conception, design, data acquisition, analysis, interpretation, software creation and drafting of the manuscript; JEC contributed to the conception, design, interpretation and revision of the manuscript. ATA contributed to the analysis, interpretation, software creation and revision of the manuscript. SHK contributed to the analysis, interpretation, software creation and revision of the manuscript. PMP contributed to the conception, data acquisition, interpretation, and revision of the manuscript. All authors read and approved of the final manuscript.

Funding

This work was supported by the *Agency for Healthcare Research and Quality* grant number 5R01HS027375. The funding agency played no role in the design of study, data collection, analysis, interpretation of results nor the writing of the manuscript.

Data Availability

The data used for this study were the IBM® MarketScan® Research Databases. These data are a propriety administrative healthcare claims database that are maintained by IBM Watson Health. These databases are used for numerous peer reviewed studies each year and can be obtained by researchers through an agreement with IBM Watson Health. The software tools used to conduct our analysis are available at the github repository https://github.com/aarmiller/delaySim as described in the manuscript.

Declarations

Ethics approval and consent to participate

Not Applicable.

Consent for publication

Not Applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Department of Internal Medicine, Roy J. and Lucille A. Carver College of Medicine, University of Iowa, Iowa City, IA 52242, USA
[2]Department of Biostatistics, College of Public Health, University of Iowa, Iowa City, IA 52242, USA

Received: 24 October 2022 / Accepted: 15 March 2023
Published online: 14 April 2023

References

1. Singh H, Graber ML. Improving diagnosis in Health Care–The Next imperative for Patient Safety. N Engl J Med. 2015;373(26):2493–5.
2. National Academies of Sciences E. Medicine. Improving diagnosis in health care. National Academies Press; 2015.
3. Garnacho-Montero J, Garcia-Cabrera E, Diaz-Martin A, Lepe-Jimenez JA, Iraurgi-Arcarazo P, Jimenez-Alvarez R, et al. Determinants of outcome in patients with bacteraemic pneumococcal pneumonia: importance of early adequate treatment. Scand J Infect Dis. 2010;42(3):185–92.
4. Zasowski EJ, Claeys KC, Lagnf AM, Davis SL, Rybak MJ. Time is of the essence: the impact of delayed antibiotic therapy on patient outcomes in Hospital-Onset Enterococcal Bloodstream Infections. Clin Infect diseases: official publication Infect Dis Soc Am. 2016;62(10):1242–50.
5. Graber ML. The incidence of diagnostic error in medicine. BMJ Qual Saf. 2013;22(Suppl 2):ii21–ii7.
6. Graber M. Diagnostic errors in medicine: a case of neglect. Jt Comm J Qual Patient Saf. 2005;31(2):106–13.
7. Singh H, Bradford A, Goeschel C. Operational measurement of diagnostic safety: state of the science. Diagnosis (Berl). 2021;8(1):51–65.

8. 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.
9. 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):155–66.
10. 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):647–55.
11. Sharp AL, Baecker A, Nassery N, Park S, Hassoon A, Lee MS, et al. Missed acute myocardial infarction in the emergency department-standardizing measurement of misdiagnosis-related harms using the SPADE method. Diagnosis (Berl). 2021;8(2):177–86.
12. Vermeulen MJ, Schull MJ. Missed diagnosis of subarachnoid hemorrhage in the emergency department. Stroke. 2007;38(4):1216–21.
13. Chang TP, Bery AK, Wang Z, Sebestyen K, Ko YH, Liberman AL et al. Stroke hospitalization after misdiagnosis of "benign dizziness" is lower in specialty care than general practice: a population-based cohort analysis of missed stroke using SPADE methods.Diagnosis (Berl). 2021.
14. Gandhi TK, Kachalia A, Thomas EJ, Puopolo AL, Yoon C, Brennan TA, et al. Missed and delayed diagnoses in the ambulatory setting: a study of closed malpractice claims. Ann Intern Med. 2006;145(7):488–96.
15. Singh H, Meyer AN, Thomas EJ. The frequency of diagnostic errors in outpatient care: estimations from three large observational studies involving US adult populations. BMJ Qual Saf. 2014;23(9):727–31.
16. 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.
17. Waxman DA, Kanzaria HK, Schriger DL. Unrecognized Cardiovascular Emergencies Among Medicare Patients.JAMA Intern Med. 2018.
18. Miller AC, Arakkal AT, Koeneman S, Cavanaugh JE, Gerke AK, Hornick DB, et al. Incidence, duration and risk factors associated with delayed and missed diagnostic opportunities related to tuberculosis: a population-based longitudinal study. BMJ Open. 2021;11(2):e045605.
19. Miller AC, Koeneman SH, Arakkal AT, Cavanaugh JE, Polgreen PM. Incidence, duration, and risk factors Associated with missed Opportunities to diagnose herpes Simplex Encephalitis: a Population-Based longitudinal study. Open Forum Infect Dis. 2021;8(9):ofab400.
20. 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.
21. McDonald KM, Matesic B, Contopoulos-Ioannidis DG, Lonhart J, Schmidt E, Pineda N, et al. Patient safety strategies targeted at diagnostic errors: a systematic review. Ann Intern Med. 2013;158(5 Pt 2):381–9.
22. Singh H, Giardina TD, Forjuoh SN, Reis MD, Kosmach S, Khan MM, et al. Electronic health record-based surveillance of diagnostic errors in primary care. BMJ Qual Saf. 2012;21(2):93–100.
23. Benedict K, Kobayashi M, Garg S, Chiller T, Jackson BR. Symptoms in blastomycosis, coccidioidomycosis, and histoplasmosis versus other respiratory illnesses in commercially insured adult outpatients, United States, 2016–2017. Clinical infectious diseases: an official publication of the Infectious Diseases Society of America; 2020.
24. Miller AC, Arakkal AT, Koeneman SH, Cavanaugh JE, Thompson GR, Baddley JW, et al. Frequency and duration of, and risk factors for, Diagnostic Delays Associated with Histoplasmosis. J Fungi. 2022;8(5):438.
25. Erickson BA, Miller AC, Warner HL, Drobish JN, Koeneman SH, Cavanaugh JE, et al. Understanding the Prodromal Period of Necrotizing Soft tissue infections of the Genitalia (Fournier's Gangrene) and the incidence, duration, and risk factors Associated with potential missed Opportunities for an earlier diagnosis: a Population-based Longitudinal Study. J Urol. 2022;208(6):1259–67.
26. Storla DG, Yimer S, Bjune GA. A systematic review of delay in the diagnosis and treatment of tuberculosis. BMC Public Health. 2008;8:15.
27. Page ES. A test for a change in a parameter occurring at an unknown point. Biometrika. 1955;42(3–4):523–7.
28. Schor S, Behar S, Modan B, Barell V, Drory J, Kariv I. Disposition of presumed coronary patients from an emergency room. A follow-up study. JAMA. 1976;236(8):941–3.
29. Pope JH, Aufderheide TP, Ruthazer R, Woolard RH, Feldman JA, Beshansky JR, et al. Missed diagnoses of acute cardiac ischemia in the emergency department. N Engl J Med. 2000;342(16):1163–70.
30. McCarthy BD, Beshansky JR, D'Agostino RB, Selker HP. Missed diagnoses of acute myocardial infarction in the emergency department: results from a multicenter study. Ann Emerg Med. 1993;22(3):579–82.
31. Lee TH, Rouan GW, Weisberg MC, Brand DA, Acampora D, Stasiulewicz C, et al. Clinical characteristics and natural history of patients with acute myocardial infarction sent home from the emergency room. Am J Cardiol. 1987;60(4):219–24.

Publisher's Note