Journal of Biomedical Informatics 85 (2018) 21–29

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

When to re-order laboratory tests? Learning laboratory test shelf-life

Gal Levy-Fix[a,*], Sharon Lipsky Gorman[a], Jorge L. Sepulveda[b], Noémie Elhadad[a]

[a] Department of Biomedical Informatics, Columbia University, 622 W. 168th Street, New York, NY, USA
[b] Department of Pathology and Cell Biology, Columbia University, 630 W. 168th Street, New York, NY, USA

ARTICLE INFO

Keywords:
Electronic health record
Ordering patterns
Data science

ABSTRACT

Most laboratory results are valid for only a certain time period (laboratory tests shelf-life), after which they are outdated and the test needs to be re-administered. Currently, laboratory test shelf-lives are not centrally available anywhere but the implicit knowledge of doctors. In this work we propose an automated method to learn laboratory test-specific shelf-life by identifying prevalent laboratory test order patterns in electronic health records. The resulting shelf-lives performed well in the evaluation of internal validity, clinical interpretability, and external validity.

1. Introduction

A common question to arise in clinical practice is *when is it time to re-order a laboratory test?* This question points to the time for which a test result remains valid for. We refer to this duration as the laboratory test *shelf-life*. Ideally, a laboratory test will be re-ordered when the previous result is considered to be no longer valid. The shelf-life of a laboratory test result changes by the clinical purpose the laboratory test is ordered for. For instance, when classifying acute renal failure, levels of creatinine is advised to be tested at specific time intervals. For suspicion of risk, injury, or renal failure the test would be re-ordered after 6 hours, 12 hours, and 24 hours (respectively) and thus have a shelf-life of less than one day [1]. When monitoring for persistent acute renal failure or loss of kidney function creatinine values remain valid for longer, warranting less frequent re-testing of 4 weeks and 3 months [1]. Thus, creatinine is an example of a laboratory test that is likely to have multiple shelf-lives, depending on the clinical context of the order. Other laboratory tests, used for primarily one clinical purpose will have a single shelf-life.

Knowing when to retest laboratory tests is often based on implicit knowledge of the physician but is rarely explicitly stated anywhere. While some laboratory test order recommendations are scattered across different manually curated guidelines, evidence based recommendations are often not readily available or adhered to [2]. If data driven laboratory test shelf-lives were available in a centralized and computable form they could be leveraged for various applications like decision support tools and for identifying general clinical practice patterns, change in practice, and misuse. For example, laboratory results available in the patient record could have an indicator pointing to whether according to the laboratory test shelf-life the result is still valid or not. This may reduce over-ordering of laboratory tests and can reinforce the clinician's reliance on available results, which is important for everyday practice. Overordering of laboratory test, while previous results are still valid, can have multiple negative effects including increased costs [3], causing unnecessary discomfort to patients [4], and an increased chance of false-positive results [5].

Laboratory test shelf-life information could also be used to inform other computational tasks using laboratory test data. For instance in the case of predictive modeling, laboratory test shelf-lives can be used to determine the informational utility of any test results observed in the data based on the last time it was tested. A certain laboratory test result from 9 months ago may not carry the same meaning or significance of a result from 2 months ago. However, one cannot just assume that the more recent the result is the better. It has been shown that in the case of blood glucose the most recent measurement is not the best predictor for the current blood glucose level [6].

In this paper we investigate a new data driven approach to automatically learn test-specific shelf-lives using the test-order patterns in the electronic health record (EHR) data of large number of patients from a single institutions. In effect, we aim to leverage the collective practice of many clinicians regarding the effective shelf-life of a laboratory test hidden in laboratory test order patterns. In this work we hypothesize that (i) time gaps between consecutive laboratory test orders in patient EHR data point to laboratory test shelf-lives and that (ii) anomaly detection method can be used to identify gap lengths that point to laboratory test shelf-lives. We test these hypotheses by implementing our proposed method on EHR data from New York-Presbyterian Hospital (NYPH) and evaluate our findings using three

* Corresponding author.
E-mail address: gf2308@cumc.columbia.edu (G. Levy-Fix).

https://doi.org/10.1016/j.jbi.2018.07.019
Received 23 February 2018; Received in revised form 15 June 2018; Accepted 19 July 2018
Available online 20 July 2018
1532-0464/

evaluation tasks. First, we evaluate the method's ability to identify common time gap lengths between laboratory test orders in the data, we then assess the clinical interpretability of the identified gap lengths, and finally we test the utility of the learned gaps in identifying phenotypes using EHR data.

Laboratory order patterns in EHR data have been analyzed before for various applications. The occurrence of laboratory test orders have been used to identify phenotypes using unsupervised probabilistic mixed membership model [7]. Occurrences of laboratory test orders have also been used to learn automated clinical order sets using Latent Dirichlet allocation (LDA) topic modeling as an unsupervised method to generate a clinical care recommendation system [8,9]. Time between consecutive laboratory test orders for the same patient have been aggregated over many patients and used to identify adherence to ordering guidelines and emerging new practices [3]. Common time lapses between repeated laboratory test orders for patients have been shown to point to the context in which laboratory tests were ordered [10]. Laboratory-context was found to help differentiate between diseases in ways laboratory test value alone could not [10]. Other works have leveraged data driven methods to identify normal laboratory test values [11] and informational needs of EHR users [12].

2. Materials and methods

2.1. Overview of approach

In this work we set out to learn laboratory test shelf-lives by identifying laboratory test order patterns in the longitudinal EHR data of many patients. This was done by calculating the time gap between consecutive laboratory orders for each patient and laboratory type. The appearance of each time gap length between laboratory orders of the same type (e.g. blood glucose test) was then counted over all patients to generate counts per laboratory test type at each gap length. As seen in Fig. 1, when the counts of time gaps between consecutive orders are aggregated over many patients they exhibit peaks at certain gap lengths. The count peaks highlighted with dotted lines at 1, 90, and 120 days seem to signal important time-to-repeat of this laboratory test, potentially its shelf-lives. Hence we sought a computational and data-driven method to identify the gap lengths for which there were significant peaks in the laboratory-gap counts. Since different laboratory tests measure different physiological markers, their shelf-lives are bound to differ. Thus we choose to analyze each laboratory test separately. To make identified peaks robust and clinically relevant we were interested in identifying laboratory test gaps equal or less than 365 days which were exhibited in the records of enough patients. As few laboratory test results are likely to have a shelf-life length longer than 1 year long.

2.2. Gap-length detection

A naive way of identifying peaks or jumps in the laboratory test gap length counts would be to set a rule-based threshold. However, one issue with this approach is that some laboratory tests are more common than others, and thus different laboratory tests may require different thresholds. Furthermore, laboratory tests that have multiple clinical purposes may have more than one shelf-life, making it hard to set one appropriate threshold. These challenges motivated us to investigate methods used in time series analysis, where a common computational task is to identify significant time points or 'events' in the data. To equate each laboratory-gap count data to a time series, gap lengths are treated as the time scale and the counts at each gap length treated as the observed random variable.

Methods commonly used to identify events in time series data include change point analysis and anomaly detection. In the context of EHR data, change point analysis has been used to detect subtle changes in emergency department admissions due to influenza-like illness [13] as well as to study the effect of an intervention on wrong-patient orders [14]. While change point analysis does aim to identify when changes have occurred in an observed series, it often assumes that the change is sustained after the change point [15]. This was not necessarily the case in the laboratory test gap-count data. By contrast, peaks in the laboratory test gap-count data seemed to more closely resemble the characteristics of anomalies in the data. Anomalies, also referred to as outliers, are often characterized as patterns in the data that do not conform to the norm [16]. That is the way the peaks in the gap counts look like, unusually high values. In the context of our study the peaks represent the exact opposite, which is frequent test order gaps. In clinical domain, anomaly detection analysis has been used for many applications including to detect unusual hospitalization patterns [17], early detection of acute infection [18], and identification of Clinical Decision Support malfunctions [19]. Other applications for anomaly detection, not restricted to time-series analysis, have included cyber-attack detection, fraud detection, industrial damages detection, image processing, textual anomaly detection, and sensor networks [16].

2.2.1. Anomaly detection

The anomaly detection method was implemented on a single type of laboratory test at a time. The general objective of anomaly detection methods is to identify unusual values in the data. In our application, the method is meant to identify frequently occurring times between the same laboratory test type, referred to as gap lengths. The gap length of an identified anomaly may signal a shelf-life of the respective laboratory test. One important aspect of the data that needed to be captured by the anomaly detection method was that gap length counts could be characterized as anomalies relative to their adjacent gap lengths but not relative to all counts. For instance for some laboratory tests, counts of

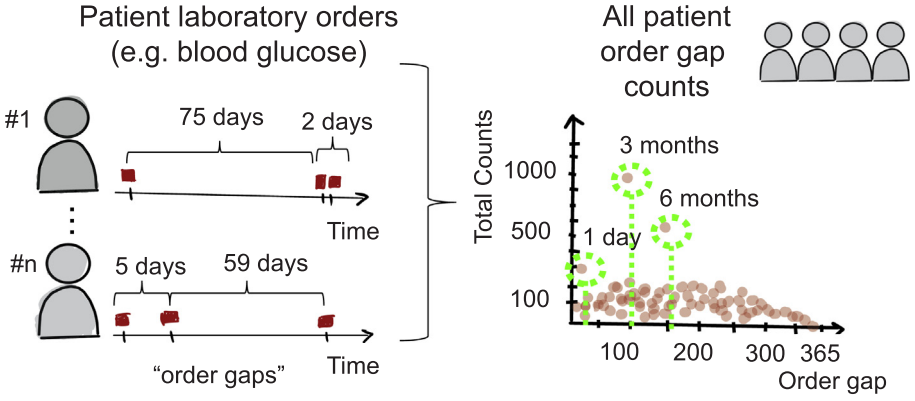

Fig. 1. Illustration of the method approach to identify prevalent laboratory test order gap lengths for a specific laboratory test (e.g. blood glucose). In the left figure the time between consecutive orders of blood glucose orders is calculated for each patient. The gap length prevalence for blood glucose is then counted across all patients to form the series seen in the right figure. Anomaly detection method is implemented on the series and identifies gap lengths with anomalously high total counts. Gap lengths longer than a year are disregarded. Given all patients' blood glucose measurement gaps, the method finds gaps at about 1 day, 3 months and 6 months; consistent with clinical guidelines for blood glucose testing.