



# An unsupervised learning method to identify reference intervals from a clinical database



Sarah Poole<sup>a,\*</sup>, Lee Frederick Schroeder<sup>b</sup>, Nigam Shah<sup>a</sup>

<sup>a</sup> Center for Biomedical Informatics Research, Stanford University, Stanford, CA, United States

<sup>b</sup> Department of Pathology, University of Michigan School of Medicine, Ann Arbor, MI, United States

## ARTICLE INFO

### Article history:

Received 18 July 2015

Revised 8 December 2015

Accepted 13 December 2015

Available online 19 December 2015

### Keywords:

Laboratory tests

Reference intervals

Unsupervised learning

Electronic health record

## ABSTRACT

Reference intervals are critical for the interpretation of laboratory results. The development of reference intervals using traditional methods is time consuming and costly. An alternative approach, known as an *a posteriori* method, requires an expert to enumerate diagnoses and procedures that can affect the measurement of interest. We develop a method, LIMIT, to use laboratory test results from a clinical database to identify ICD9 codes that are associated with extreme laboratory results, thus automating the *a posteriori* method. LIMIT was developed using sodium serum levels, and validated using potassium serum levels, both tests for which harmonized reference intervals already exist. To test LIMIT, reference intervals for total hemoglobin in whole blood were learned, and were compared with the hemoglobin reference intervals found using an existing *a posteriori* approach. In addition, prescription of iron supplements were used to identify individuals whose hemoglobin levels were low enough for a clinician to choose to take action. This prescription data indicating clinical action was then used to estimate the validity of the hemoglobin reference interval sets. Results show that LIMIT produces usable reference intervals for sodium, potassium and hemoglobin laboratory tests. The hemoglobin intervals produced using the data driven approaches consistently had higher positive predictive value and specificity in predicting an iron supplement prescription than the existing intervals. LIMIT represents a fast and inexpensive solution for calculating reference intervals, and shows that it is possible to use laboratory results and coded diagnoses to learn laboratory test reference intervals from clinical data warehouses.

© 2015 Elsevier Inc. All rights reserved.

## 1. Introduction

Reference intervals are a critical part of the interpretation of laboratory results, and have been described as the most widely used medical decision-making tool [1]. Ideally, a patient specific approach would be taken for the interpretation of laboratory results, because historic data from the patient of interest is the best indicator of the normality of the current results [2]. Such a personalized approach would allow small fluctuations to be identified, a desirable characteristic given recent studies showing the predictive power of gradual temporal trends [3,4]. However, large amounts of patient specific data are rarely available. For this reason, population based reference intervals are the norm, and their use for result interpretation is ubiquitous in laboratories today [2].

The concept of reference intervals was first widely adopted in 1969 [5], and today reference intervals are required for all laboratory developed tests [6]. Aytekin and Emerk provide three reasons

that the description and definition of reference intervals is not a solved problem [7]. First, as new biomarkers are identified and adopted in clinical practice, it is important that the associated reference intervals are defined. For example, when identifying new protein biomarkers, reference intervals for specific populations can quantify the effect of age, sex, ethnicity or race on protein concentration, ensuring that these effects are considered in interpreting the study results [8]. Second, a large number of reference intervals that are currently in use were established for a Caucasian population, and thus are not always suitable for use in interpretation of laboratory results of patients from different ethnic groups. Separate reference intervals should be established for different races and ethnicities. Third, many of the reference intervals that are in use today were developed many years ago, and are specific to measurement techniques and instruments that are no longer in use. These reference intervals should be updated, or validated for today's measurement techniques and instrumentation.

Determining reference intervals requires a group of clinically normal individuals in order to obtain a large number of normal results for the test of interest [9]. This process is costly,

\* Corresponding author.

E-mail address: [spoole@stanford.edu](mailto:spoole@stanford.edu) (S. Poole).

time-consuming and complex [10–13]. The rules used to determine patient ‘normality’ vary widely. Particularly rigorous criteria may lead to only a fraction of healthy people being accepted into the test group [2]. Therefore, the initial pool of possible subjects must be large, increasing the cost of the process. A group of at least 120 reference individuals is typically recommended to establish reference intervals [14]. It is recommended that all laboratories develop their own reference intervals [2,12], to ensure that resulting intervals are appropriate for the population being served [10,12,13,15].

For most clinics, the development of their own reference intervals is a ‘practical impossibility’ [13], and reference intervals are regularly adopted from the literature or other sources [11,12]. Reference intervals from manufacturers’ guidelines or published literature may be used for FDA approved test sets, or in situations where the specimens are difficult to obtain [9]. When a population based reference interval is adopted, the population used for the development of reference intervals should be as close as possible to the patient group being served, with the exception of the disease being tested for [15]. The International Federation of Clinical Chemistry and Laboratory Medicine (IFCC) has published guidelines for the establishment and use of laboratory reference intervals [14,16].

When not establishing a new reference interval and instead adopting an already established reference interval, the IFCC recommends that each laboratory verify the appropriateness of the reference intervals for their population [17]. This verification is required by law for FDA approved or cleared devices [6], and it is recommended that a group of at least 20 reference individuals are included [14]. The verification process is more affordable than establishing reference intervals, since the size of the reference group required is much smaller, however the cost involved is still a disincentive for laboratories. Healthy laboratory staff are often used in the verification process. Reference intervals based on tests from laboratory staff give a biased result that is not representative of the patient population. Ideally, a patient’s comorbidities and characteristics should be mirrored in the reference population [18,19]. For example, results from hospitalized patients should not be interpreted using reference intervals developed from healthy ambulatory populations [2], and patient traits such as BMI and nutritional habits (e.g. vegetarian diet) should match the reference population [13]. Such subdivision of the population based on patient characteristics will decrease the inter-individual variability, and tighten the reference intervals [13]. Creating reference intervals based on such subdivision was found to decrease the number of patients who had unexplained thrombocytopenia [20], showing the importance and impact of such personalization. It would be helpful to have reference intervals would require a scalable method, which is not currently available.

The use of laboratory data to develop reference intervals is generally via an indirect approach, in which the reference population is chosen based on some statistical criteria [10–12,15,21]. For example, a basic method may classify any results more than two standard deviations from the mean as outliers, while a more complex method may model the results as a sum of two Gaussian distributions: one from the outlier results, and one from the non-outlier results [11]. These indirect approaches are often used to calculate reference intervals for pediatric and geriatric populations, where data is sparse [22,23]. The IFCC recommends that the characteristics of the reference population are clearly defined [13], which is not achievable using these purely statistical approaches. Therefore, the IFCC does not recommend such indirect approaches.

With the increasing availability of clinical data, the reference population can be defined based on review of the electronic medical records in an *a posteriori* approach [13], which meets the IFCC

recommendations but is still costly and time consuming due to the time required from experts. For each test of interest, an expert must define the diagnoses and procedures associated with abnormal results. Patients with the corresponding ICD9 and CPT codes must then be excluded from the cohort before the reference intervals are calculated. This *a posteriori* process has been shown to give accurate results [2,12,21], and is currently the most commonly used system for defining reference intervals from data, with the Australasian Association of Biochemists (AACB) suggesting the possible use of this method at an individual laboratory level [24]. However, this approach does not overcome the problem of scalability, as exclusion rules must be developed manually for each test and patient population of interest.

There have been attempts to automatically identify and remove diseased patients from the data. Grossi et al. [25] work under the assumption that repeated tests correspond to a higher probability of disease, and select patients that have only one measurement for the laboratory test of interest. Kouri et al. [2] exclude all but the first measurement for each patient in order to avoid a potential bias from repeated measurements. Both of these methods drastically decrease the size of the patient population making it difficult to maintain statistical power and to draw valid conclusions. Weber and Kohane [26] investigate laboratory test repeat interval as a way to assess normality of a test result. They found that the repeat interval between tests was shorter when the initial test was abnormal, and were able to use this to determine which test results physicians consider to be normal. While this method has good results, the results being excluded are not directly tied to patient characteristics, meaning that the characteristics of the remaining patient population are not defined, and this process is not in line with IFCC recommendations.

We propose that it is possible to use laboratory results and coded diagnoses to learn laboratory test reference intervals. We present an automated method to learn the ICD9 codes that are significantly associated with extreme laboratory results using existing clinical data. Doing so removes the key bottleneck—the need for expert definition of exclusion criteria—of the *a posteriori* method, while still meeting the IFCC recommendations. The crux of our method is an unsupervised mechanism to identify the ICD9 codes associated with extreme laboratory results (Fig. 1). Patients with these codes are then removed from the reference population in an iterative manner. The automated nature of this process means that reference intervals can be inexpensively calculated for many different patient populations, and that these calculations can be repeated regularly to ensure accuracy. The ability to quickly and inexpensively find reference intervals using different sets of results will help to address the problem of biases in laboratory test data. At its core, our method builds on the observation that a significant proportion of specimens that are sent for laboratory testing are normal [11]. This method is named LIMIT – Laboratory Information Mining for Individualized Thresholds.

## 2. Material and methods

LIMIT consists of 7 steps, from data extraction to comparison of the resulting reference intervals, as shown in Fig. 1. Laboratory results, ICD9 codes and patient demographics are extracted from STRIDE (step 1), and the patients are split into two groups based on Hampel outlier detection of their laboratory results: those with no outlier results, and those with at least one outlier result (step 2). Fisher’s Exact Test is then used to find ICD9 codes that are overrepresented in the outlier group (step 3). The ICD9 code that has the lowest *p*-value is identified, and if the *p*-value is below a chosen cutoff, this code is considered to be associated with extreme laboratory levels (step 4). All patients with any mention of this ICD9

Download English Version:

<https://daneshyari.com/en/article/6927905>

Download Persian Version:

<https://daneshyari.com/article/6927905>

[Daneshyari.com](https://daneshyari.com)