Clinical Biochemistry 46 (2013) 1220–1227

Contents lists available at ScienceDirect

Clinical Biochemistry

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

Review

A systematic review of statistical methods used in constructing pediatric reference intervals

Caitlin H. Daly [a,b], Xiaofeng Liu [a,c], Vijay L. Grey [c], Jemila S. Hamid [a,b,c,*]

[a] Clinical Epidemiology and Biostatistics, McMaster University, Canada
[b] Mathematics and Statistics, McMaster University, Canada
[c] Pathology and Molecular Medicine, McMaster University, Canada

ARTICLE INFO

Article history:
Received 21 April 2013
Received in revised form 13 May 2013
Accepted 15 May 2013
Available online 29 May 2013

Keywords:
Pediatric
Reference interval
Reference limit
Reference value
Statistical methods

ABSTRACT

Objectives: This study aims to investigate current medical literature with focus on statistical methods used to construct pediatric reference intervals and identify potential gaps within the process of reference interval estimation.

Design and methods: A systematic review of methods was performed. Extensive search criteria were developed and search was conducted on Embase, Medline, and PubMed databases to identify relevant articles. The articles were further screened using predefined inclusion and exclusion criteria. The selected articles were then included in our final systematic review.

Results: Our review reveals that there are gaps within current methodology and reporting of pediatric reference intervals. Not all publications followed the Clinical and Laboratory Standards Institute (CLSI) guidelines, and there is a large variation in the methods used. Discrepancies particularly arose when reference intervals were calculated for partitions with small sample sizes. In addition, the dynamic nature of pediatric data was not mostly captured when certain partitioning techniques were used.

Conclusions: There are areas within the pediatric reference interval development process that need attention. Partitioning methods particularly need to be explored with the goals of reducing subjectivity and enabling researchers to capture the best representative partitions possible. Moreover, the complicated nature of pediatric data often limits the sample size available for each partition and appropriate methods need to be considered in such cases. Researchers are also strongly encouraged to accompany their reference limits with confidence intervals to show sampling variability and demonstrate precision of their limits. These issues exemplify the need for a pediatric specific guideline that outlines a standardized way of establishing reference intervals.

Contents

Introduction

Reference intervals provide clinicians a normal range for comparison when evaluating and interpreting a patient's laboratory test results and are one of the most important tools in diagnostic and laboratory medicine. Current guidelines define reference intervals as ranges of values within which a specified percentage of measurements from healthy individuals would fall [1]. A wide variety of factors affect the validity of reference intervals. These include reference population,

Abbreviations: BMI, Body Mass Index; CALIPER, Canadian Laboratory Initiative in Pediatric Reference Intervals; CHILDx, Children's Health Improvement Through Laboratory Diagnostics; CLSI, Clinical and Laboratory Standards Institute; KiGGS, German Health Interview and Examination Survey for Children and Adolescents; LOOK, Lifestyle Of Our Kids; REB, Research Ethics Board.

* Corresponding author at: Clinical Epidemiology and Biostatistics, 1280 Main Street West, HSC-2N35B, McMaster University, Hamilton, Ontario L8S 4K1, Canada.
E-mail address: jhamid@mcmaster.ca (J.S. Hamid).

0009-9120/$ – see front matter
http://dx.doi.org/10.1016/j.clinbiochem.2013.05.058

sampling strategy, sample size, analytical factors such as instrumentation, gender, age, and other demographic and lifestyle factors. The statistical methods used to construct reference intervals also play a major role on the resulting intervals. However, they are often overlooked as factors that might have a considerable effect on the validity of the reference intervals.

Several statistical methods are available to estimate reference intervals. The parametric approach is one of the most commonly used, where data is assumed to follow a Gaussian distribution [2]. On the other hand, the non-parametric approach does not assume any distribution for the data and typically uses the central 95% range of the observed data to provide reference intervals [1]. A more complicated robust approach, which uses an iterative process based on the median and median absolute deviation of the observed data is also available [3].

Due to the lack of a standardized statistical approach in selecting which methods to use, the resulting variability in the development of previously published reference intervals raises questions. The Clinical and Laboratory Standards Institute (CLSI) has recognized the need to establish reference intervals through a more systematic means and thus has provided a reasonable guideline to resolve this issue [1]. This guideline suggests the non-parametric approach using a minimum of 120 healthy samples to establish reference intervals for a particular reference group [1]. In practice, however, obtaining 120 samples is very expensive and labor intensive, and at times impossible. In fact, a study, published by the College of American Pathologists, showed that among 129 reference intervals for potassium constructed by in-house laboratories, 60 of them (46.5%) used less than 50 samples, 28 (21.7%) used less than 100 samples, and 32 (24.8%) used more than 100 samples [4,5]. The robust method is recommended as an alternative statistical approach when the collection of 120 samples is not possible. Nevertheless, the majority of studies often use the non-parametric approach regardless of sample size, perhaps because of the complexity of implementing the robust approach.

The CLSI guideline is primarily developed for the general population and fails to consider factors unique to pediatric populations. Children are continuously growing from birth to adolescence, and hence more age partitions (than adult populations) are often required. In addition to the standard covariates for adult populations (e.g., age, body mass index (BMI), gender, or ethnicity), maturity markers such as Tanner stage can greatly influence the composition of pediatric populations for which reference intervals should be provided. To capture all of these factors and ensure reference intervals are applicable, several partitions are warranted in pediatric data. The first year of life, for example, requires many partitions to highlight extrauterine adaptation and development patterns [6].

Another important consideration when establishing pediatric reference intervals is achieving sufficient sample size for every partition. This is particularly challenging in pediatric populations. Children are smaller than adults and thus blood procurement can be difficult. For example, 10 mL of blood could constitute to 10% of blood volume in a baby [7]. In addition to Research Ethics Board (REB) constraints, parental consents and costs make this a very demanding task.

Several national projects such as Canadian Laboratory Initiative in Pediatric Reference Intervals (CALIPER) [8], Children's Health Improvement Through Laboratory Diagnostics (CHILDx) [9], German Health Interview and Examination Survey for Children and Adolescents (KiGGS) [10], and Lifestyle Of Our Kids (LOOK) [11] are currently underway to address the issue of outdated and unreliable pediatric reference intervals published in the past. With these projects underway, it is very important now more than ever to develop an outline of circumstantial methods in order to avoid the unnecessary variability that may exist between these groups' published intervals, strictly due to differences in statistical methodology. Minimizing the differences between methodologies and providing a unified framework for selecting appropriate statistical methods will help clinicians compare reference intervals that are produced by various studies and identify any differences that may exist between populations. To make this possible, a thorough investigation is required to determine the impact various statistical methods have on resulting reference intervals. However, it is of great importance to first assess current practice and identify the different statistical methods that are available for establishing pediatric reference intervals.

This systematic review was designed with the aim of 1) investigating current literature on pediatric reference intervals with a focus on statistical methods that are used to construct pediatric-specific reference intervals, 2) identifying gaps in the choice and implementation of the methods and reporting of the results. This will allow researchers to gain some insight into weak areas of current practice and provide direction when developing pediatric reference intervals. We hope this review will lay the groundwork for performing comparisons of existing and emerging methods under various scenarios (sample size, distributions, partitions, and outliers). We believe that these comparisons will ultimately lead to standardized approaches and establishment of guidelines specific to pediatric populations.

Material and methods

An electronic search on the Embase, MEDLINE and PubMed databases was conducted on May 28, 2012. We pre-identified three themes as our search criteria. These themes were: “establishing”, “pediatric”, and “reference intervals”. Within each theme, we developed a list of keywords or phrases, which included various synonyms of the three themes commonly used in past literature. Effort was made in the search to ensure that some combination of the three themes was necessary for an article to be included in the search results. Search terms within each theme were combined with “OR”, and themes were combined with “AND”. Fig. 1 represents the three themes as well as the search words within each theme that were used in our search of literature.

English articles published from January 1st, 2011 to the present were considered in this systematic review. Duplicated articles were deleted before the initial screening process. After removing duplicates, two authors (CD and XL) were then presented with the resulting unique articles. The authors proceeded to independently review the title and abstracts of these articles against predetermined inclusion and exclusion criteria. Articles included in the systematic review presented new pediatric reference intervals established by the authors for intended public or in-house use. Articles were excluded if the reference intervals were calculated based on unhealthy samples, samples were outside the birth to less than 19-year age range, and/or samples were from a non-human population. In addition, if a study in the article used longitudinal data, strictly cited reference intervals for diagnostic or validation purposes, or simply did not establish new reference intervals, it was excluded. These restrictions of our review may have resulted in the exclusion of possibly valuable studies. In particular, recent articles from the well-known German KiGGS study, written in German, as well as the longitudinal Australian LOOK study, were not included in this review [12,13]. Following the exclusion of irrelevant articles, the same two authors reviewed the resulting articles' full text against the inclusion and exclusion criteria. If any disagreement arose during these processes, it was resolved either by discussion or consultation through a third author (JH). After reviewing all prospective articles, the reference lists of the included full-text articles were screened for additional relevant articles.

Results

In total, 373 articles were initially returned from the keyword search, of which 195 were found to be unique. Through the initial screening of titles and abstracts, using the inclusion/exclusion criteria outlined in the Materials and methods section, 37 articles were