Estimation of Sex-Age Specific Clinical Reference Ranges by Nonlinear Optimization Method

Takeo Shibata a, Yoichi Ogushi a, Teppei Ogawa b, Takashi Kanno c

a Tokai University, Kanagawa, Japan
b Joto Hospital, Yamanashi, Japan
c Hamamatsu Medical Corporation, Shizuoka, Japan

Abstract

Most reference ranges are not considered sex-age specific differences. We collected about 700,000 health examination data with 24 items to estimate sex-age specific clinical reference ranges. We proposed nonlinear optimization method as a new method compatible with NCCLS guideline. All items showed sex-age specific differences of reference ranges. Especially, hepatic functions in young aged men and all aged women, diabetic functions in young aged people, blood pressures in older people, and total cholesterol in all aged people might have serious problems. Some abnormal individuals might not be detected using established reference ranges, on the other hand, some normal individuals might be treated excessively.

Keywords:
Sex-age specific clinical reference ranges; Nonlinear optimization method; NCCLS guideline

1. Introduction

Recently, sex-age specific differences in medicine have been attracted attentions. But most reference ranges are not considered sex-age specific differences. Many guidelines defined reference ranges by the results based on patient’s data [1,2]. To determine health examination values, the reference ranges should be defined based on pure normal individuals. So we collected about 700,000 health examination data to estimate sex-age specific reference ranges.

NCCLS (National Committee for Clinical Laboratory Standards) developed a guideline to define and determine reference ranges for clinical laboratory data in 2000 [3]. It would be a standard method to determine reference ranges, because the basis of the guideline is compliant with the committee of IFCC (International Federation of Clinical Chemistry) and ICSH (International Council for Standardization in Haematology). This guideline includes the methodological approaches and the procedure which is recommended to establish reference ranges.

But it is hard to apply NCCLS guideline for a determination of clinical reference ranges, because in-depth exclusion criteria and many partitioning factors are required, and selected individuals might be less than 5% of original individuals.

So we proposed a new method to establish sex-age specific clinical reference ranges which found a normal distribution mathematically. Our new method does not require any exclusion
criteria and partitioning factors, and most individuals are used in a calculation of reference ranges. A confirmation of a compatibility with NCCLS method is require to verify a validity of our methods.

2. Materials

2.1. Data collection

Health examination data was collected between April 1 2002 and March 31 2003 with 45 institutions retrospectively in Japan. 45 institutions participated our study, and about 700,000 cases were collected. 24 items were collected for each institution (Table 1).

<table>
<thead>
<tr>
<th>BMI(upper &amp; lower)</th>
<th>Systolic Blood Pressure</th>
<th>Diastolic Blood Pressure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Protein</td>
<td>Albumin</td>
<td>G-GTP</td>
</tr>
<tr>
<td>AST(GOT)</td>
<td>ALT(GPT)</td>
<td>Total Bilirubin</td>
</tr>
<tr>
<td>LDH</td>
<td>Alkaline Phosphatase</td>
<td></td>
</tr>
<tr>
<td>TC(upper &amp; lower)</td>
<td>Triglyceride</td>
<td>HDL cholesterol</td>
</tr>
<tr>
<td>Fasting Blood</td>
<td>HbA1c</td>
<td>LDL cholesterol</td>
</tr>
<tr>
<td>Glucose</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Creatinine</td>
<td>Uric Acid</td>
<td></td>
</tr>
<tr>
<td>Red Blood Cell</td>
<td>Hemoglobin</td>
<td>Hematocrit</td>
</tr>
<tr>
<td>White Blood Cell</td>
<td>Platelet(upper &amp; lower)</td>
<td></td>
</tr>
</tbody>
</table>

Measuring methods had been standardized by JSCC (Japan Society of Clinical Chemistry), and all items could be categorized by one or two methods. Cache’ ver.5.0 was used to make database and to calculate reference ranges.

2.2. Data integration

The data were separated by measuring methods, sex and each 5 years old between 20 and 79 years old. At first, reference ranges were calculated for men aged between 50 and 54 by each institution to confirm if the measuring methods were right, because that age group had the biggest number of individuals.

To integrate all data, a random sampling technique was used to reduce the number and make no influence for a particular institution when the sample had more than 2,000 cases.

3. Methods

3.1. NCCLS method

NCCLS (National Committee of Clinical Laboratory Standard) provided a guideline to define and determine reference intervals in the clinical laboratory in June 2000. NCCLS guideline defines an exclusion criteria and partitioning factors to select reference individuals. To calculate reference ranges being compliant with NCCLS guideline, in-depth data with unified formats is required. It is hard to collect those exclusion criteria and partitioning factors.
3.2. Normal distribution test

Reference ranges are calculated as 95% confidence intervals of a normal distribution. If the distribution is not shown a normal distribution, the clinical laboratory test data is generally transformed as log, square, third power, square root, and third root. Chi-square test was used to test a normal distribution. At first, histograms were drawn. Then the expected value as a normal distribution was calculated by each histogram. Chi-square test compared the number of individuals and the expected value. If a distribution of original data is not shown as a normal distribution, the data which has the smallest p value of chi-square test is selected.

If a distribution without any transformations shows a normal distribution, the original data is used to calculate the reference ranges. Then 2.5 percentile and 97.5 percentile are used as the reference interval.


In this study, NCCLS method was used to estimate the reference ranges first. But reasonable results were not estimated, because the data of exclusion criteria and partitioning factors were not enough to refine a pure normal group. So a nonlinear optimization method was used to estimate clinical reference ranges. The procedure of this method is as follows; (1) separate data by sex and each 5 years old in original data (without any transformations), (2) search linear areas on Q-Q plot, (3) make fitting a normal distribution to histogram between the linear areas data (4) calculate 95%CI of the fitted normal distribution.

To confirm the compatibility with NCCLS method, we used the data of Tokai University Health Checkup Center, because they contain detailed questionnaires. After separation of the data by sex and age, we selected the reference individuals by questionnaire data based on NCCLS method. Then we examined if the distribution shows a normal distribution. We compared histograms and 95%CIs between NCCLS method and those by our new method. Results of hepatic function tests were used to compare these two methods, because they had high abnormal rates.

On the other hand, we performed simulation of random walks with recovery force to confirm that a pure normal group shows a normal distribution. To examine the influence of abnormal rates, artificial data which had abnormal rates from 0 to 40 percentages were generated by a boot-strap method. Then the reference ranges calculated by the new method were compared.

3.3.1. Searching a linear area on Q-Q plot

Figure 1 shows a Q-Q (quantile-quantile) plot of HbA1c in the cases with men aged between 50 and 54. If a distribution is a normal distribution, Q-Q plot shows a linear line. A linear area including median on Q-Q plot means a pure normal area. So a linear area is searched on Q-Q plot.

The procedure to search a linear area as follows; (1) calculate and draw a regression line including the point of median, (2) take the higher intersection point, (3) narrow the targeted area to the intersection point and recalculate a regression line, (4) compare the intersection point and the previous one. Then if they are not corresponded, go back to (3), and if they are corresponded, (5) determine the higher edge of the linear area as the point which has no changes with the previous point, (6) go back to (1) and take lower intersection point on the procedure (2) to determine the lower edge (Figure 2).

3.3.2. Estimation reference ranges by nonlinear optimization method

The histograms ranged by the linear area are used to estimate reference ranges. A normal distribution curve is fit the histogram by nonlinear optimization method. Figure 1 shows an
original histogram and a pure normal distribution curve. Reference ranges are calculated as 95% C.I. using the mean and the standard deviation of the pure normal distribution.

Figure 1 - Searching a linear area and nonlinear optimization method

Figure 2 – Procedure of searching a linear area on Q-Q plot

4. Results

4.1. Compatibility with NCCLS method

The AST(GOT) data of 60s men who were selected as reference individuals showed a normal distribution. And the histogram and the 95%CI by our new method were coincident with NCCLS method. Our new method without any questionnaire data can estimate sex-age specific clinical reference ranges coincident with NCCLS method.

4.2. Simulation of a pure normal distribution

A random walk simulation proved that pure normal individuals show a normal distribution where they are affected with some noises and have recovery forces.
4.3. Estimation of sex-age specific reference ranges

All items showed sex-age specific differences of reference ranges. The sex-age specific reference ranges is shown on the following URL:
http://www.mi-tokai.com/defaulte.htm or http://mi.med.u-tokai.ac.jp/defaulte.htm

Though reference ranges have been separated by sex among some items, strong dependences with age are shown. Many reference ranges were similar with the ranges of middle aged men. A menopause had a potent influence on reference ranges in women.

Hepatic functions show potent influences by sex and age (Figure 3). Total cholesterol shows a crossover by sex and a potent influence by age and sex (Figure 4). Systolic blood pressure shows a linear upward influence by age (Figure 5). HbA1c shows a linear upward influence by age, and fasting glucose also shows a potent influence by sex. Albumin shows a linear downward influence by age and a clear menopausal effect in women (Figure 5). Alkaline Phosphatase and LDH show clear menopausal effects.
4.4. Compilation of pure normal distributions and influence of abnormal rates

Though sex-age separated data selected as normal individuals showed a normal distribution, sex-age mixed data were not shown even if normal individuals were selected. The data with abnormal rate less than 40% had no problem to determine clinical reference ranges using our new method.

5. Conclusion

Nonlinear optimization method was able to estimate reference ranges without exclusion of abnormal individuals, and it has compatibility with NCCLS method. It also means that our new method is able to calculate reference ranges using many individuals. Some simulations reconfirmed the validity of our new method. So our new method is more useful than NCCLS method, because detailed exclusion criteria and partitioning factors are not necessary and more cases are effective for the calculation. It is the most important to separate by age (deservedly by sex), and many studies have been shown age specific differences of clinical laboratory data [5,6]. NCCLS method also includes age and sex as partitioning factors.

Sex-age specific reference ranges by nonlinear optimization method shows interactions between sex and age in all items. Those reference ranges should be considered with the influence with sex and age. Especially, hepatic functions in young aged men and all aged women, diabetic functions in young aged people, blood pressures in older people, and total cholesterol in all aged people might have serious problems. Some abnormal individuals might not be detected using established reference ranges, on the other hand, some normal individuals might be treated excessively.

In this study, sex-age specific and “health associated” reference ranges were established. But some items might not be used for clinical decision making, because high risk people should be made decision by the ranges for high risk group.

6. References


Address for correspondence
259-1193 Boseidai, Isehara, Kanagawa, Japan., Department of Medical Education and Informatics, School of Medicine, Tokai University, E-mail : tshibata@is.icc.u-tokai.ac.jp