Establishment of PIVKA-Ⅱ Reference Intervals From Hospital-stored Data: A Comparison Analysis

Lewei Zhou
South-Central University for Nationalities
Qiyuan Su
University of Illinois at Urbana-Champaign
Yan Yao
Renmin Hospital of Wuhan University
Meixian Xiang
South-Central University for Nationalities
Jiesheng Zhen
Renmin Hospital of Wuhan University
Hanwen Su (hanwensu@whu.edu.cn)
Renmin Hospital of Wuhan University
Zhengjun Zhang
University of Wisconsin

Research Article

Keywords: reference interval, abnormal prothrombin, gender-partitioned groups, Laboratory Information System

Posted Date: November 30th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1040511/v1

Establishment of PIVKA-II reference intervals from hospital-stored data: a comparison analysis

Lewei Zhou[1#], Qiyuan Su[2#], Yan Yao[3#], Meixian Xiang[1], Jiesheng Zhen[3], Hanwen Su[3*], Zhengjun Zhang[4*]

1, School of Pharmaceutical Sciences, South-Central University for Nationalities, Wuhan, 430074 PR China

2, Department of Statistics, University of Illinois at Urbana-Champaign, Urbana, IL 61820, USA

3, Department of Clinical Laboratory, Renmin Hospital of Wuhan University, Wuhan, 430061, PR China

4, Department of Statistics,University of Wisconsin,Madison, WI, 53706, USA

[#] These authors contributed equally to this work.

Hanwen Su, M.D

Department of Clinical Laboratory,

Renmin Hospital of Wuhan University,

Wuhan, 430061, PR China

Phone:+86 027-88041919

E-mail: hanwensu@whu.edu.cn

Zhengjun Zhang P.H.D.

Department of Statistics,

University of Wisconsin,Madison,

WI, 53706, USA

Phone: 608-262-2614

E-mail: zjz@stat.wisc.edu

Abbreviation

PIVKA-II, abnormal serum prothrombin; HCC, Hepatocellular carcinoma; LIS, Laboratory Information System; AFP, Alpha-feto-protein; CLSI, The American Clinical and Laboratory Standardization Institute; CSF, cerebrospinal fluid; CDF, cumulative distribution function; IQR, interquartile range.

Funding statement

This work was supported by the Natural Science Foundation of China (No.32072767), and the Fundamental Research Funds for Central Universities South-Central University for Nationalities (No. CZZ21011 No.3212021sycxjj382) and the Fundamental Research Funds for Health Commission of Hubei Province (ZY2021M061)

Conflict of interest disclosure

The authors declare no conflict of interest.

Ethics approval statement

The ethics committee of Renmin Hospital of Wuhan University approved the study and the informed consent was waived by the ethics committee of Renmin Hospital of Wuhan University due to the retrospective analysis test data from clinical medical laboratories, which did not involve private patient information.

Abstract

Objective The authors aimed to explore methods to establish indirect reference intervals for PIVKA-II from hospital-stored data.**Method** 7623 patient specimens of the Renmin Hospital of Wuhan University were collected. Indirect reference intervals were established based on the hospital-stored data with four different methods, including the Hoffmann method (HM), revised Hoffmann method (HMCDF), E-M algorithm-based method (EMBCT), and a recent estimator (KOSMIC). According to CLSI C28-A3 guidelines, 369 healthy specimens were collected. The authors tested the difference between reference intervals of gender-specific and age-specific subgroups using Harris and Boyd's test. Finally, the averaging result of estimates was calculated according to how likely each model was.**Results** The indirect reference intervals of PIVKA-II based on LIS data were 0 to 35.30 mAU/mL (HM), 0 to 31.48 mAU/mL (HMCDF), 0 to 30.78 mAU/mL (EMBCT), 0 to 36.17 mAU/mL (KOSMIC) and 0 to 31.48 mAU/mL (averaging) respectively, and the reference intervals based on healthy group were 0 to 32 mAU/mL. Compared with HM, EMBCT and KOSMIC, HMCDF and the averaging result was closer to those of the health group. Significant difference was detected between gender-partitioned subgroups, and the reference upper limit in the female group was smaller than the male group.**Conclusions** The authors established the indirect reference intervals of PIVKA-II for the Wuhan population, which could be used to the clinical reference intervals. The framework proposed could help clinical laboratory set their reference intervals of test items.

Keywords: reference interval, abnormal prothrombin, gender-partitioned groups,

Laboratory Information System

Introduction

Hepatocellular carcinoma (HCC) is one of the most common malignant tumours globally, with the mortality rate ranking among the top three malignant tumours of the digestive system. The annual fatality and new liver cancer cases in China account for about 50% of the total number of liver cancer cases worldwide[1]. The initial symptoms of hepatocellular carcinoma are insidious and difficult to detect at an early stage[2], making early clinical diagnosis and treatment very challenging. Alpha-feto-protein (AFP) has been the most widely used tumour marker for HCC. However, 30%-40% of HCC patients have negative serum AFP. Therefore, there is a necessary and urgent need to find new tumour biomarkers for hepatocellular carcinoma[3]. The new tumour biomarkers can effectively compensate for the AFP-negative diagnosis of hepatocellular carcinoma patients.

Human abnormal prothrombinogen (PIVKA-II), a protein induced by vitamin K deficiency or antagonist II, is also known as dextro-γ-carboxy-prothrombinogen. Published research suggests that patients may have abnormal liver metabolism due to vitamin K deficiency, allowing incomplete carboxylation of several glutamates near the amino terminus of prothrombinogen to form abnormal prothrombin and lose the related coagulation activity [4,5]. In HCC patients, the endoplasmic reticulum can not carboxylate PIVKA-II to become active normal prothrombin, resulting in elevated serum PIVKA-II levels. If PIVKA-II was found to be abnormally elevated in serum, it

might provide a basis for detecting HCC to some extent. The literature suggests that PIVKA-II is elevated in a certain percentage of HCC patients[6] and has high diagnostic specificity, especially in AFP-negative hepatocellular carcinoma patients. It has been suggested that PIVKA-II is generally more effective than AFP and has an excellent clinical application in combination with AFP in tandem or a parallel inspection[7].

Due to insufficient literature and research related to PIVKA-II in China, many clinical laboratories generally follow two ideas for determining its reference interval. One is to cite the literature or the manufacturers' reagent instructions. The second is to transfer and verify its biological reference interval. Due to race, age, and gender differences and the influence of testing methods, instruments, and reagents, the resulting reference intervals obviously cannot provide accurate clinical diagnosis and treatment guidance. They may even cause trouble to clinical work and patients.

The American Clinical and Laboratory Standardization Institute (CLSI) document C28-A3 is currently widely recognized, which recommends that each clinical laboratory establish reference intervals appropriate to its condition[8]. However, setting reference intervals following the CLSI files takes a lot of time and money. Moreover, it is even more difficult in clinical work when reference intervals are needed to differentiate age and genders, especially newborns and the elderly. Besides, for some particular analytes, such as cerebrospinal fluid (CSF), it becomes tough to establish reference intervals for tests in these situations[9].

With computer information technology development, hospitals have gradually established and improved their laboratory information systems (LIS), which store

many test data and contain a massive amount of value yet to be explored. Several literature pieces have proposed the methods and ideas of establishing reference intervals based on existing data[10]. In this study, the indirect biological reference intervals of PIVKA-II were shown with four methods, respectively, and an averaging result was presented with these four models. Based on the CLSI document guidelines, the authors established and briefly validated the reference intervals for our region's healthy population. Compared with the manufacturer's reagent instruction's reference interval, the local PIVKA-II was initially established to help transfer and validate the laboratory reference interval and reasonably guided the clinical diagnosis and treatment.

Materials and method

This was a test data retrospective analysis from clinical medical laboratories. The protocol was approved by the institutional review board at The Renmin Hospital of Wuhan University. The study was conducted in accordance with the Declaration of Helsinki and adhered to Good Clinical Practice guidelines.

Storage group

All PIVKA-II test values are stored in the Laboratory Information System (LIS) of the Renmin Hospital of Wuhan University, totaling 7623 cases, counted from January to December 2018, with ages ranging from 1 to 105 years old. Among them, 4860 patients were male, aged 1 to 105 years old; 2763 cases were female, aged 6 to 99 years old.

Health group

The health group includes 369 cases of apparently healthy individuals from the People's Hospital of Wuhan University's physical examination centre, counted from January to December 2018, ranging from 15 to 70 years old. Among them, 193 cases are male, aged 15 to 70 years old; 176 cases are female, aged 15 to 70 years old. Inclusion and exclusion criteria were strictly defined. Individuals with the liver system, benign liver disease, or malignant tumours of reproduction or gastrointestinal that could cause elevated PIVKA-II were excluded based on medical history and relevant data. After the screening, 369 cases were retained in the health group.

PIVKA-II test

3 mL of venous blood was collected in a rest fasting state using a vacuum blood collection tube containing procoagulant (Becton Dickinson and Company, BD Vacutainer, USA) and then centrifuged at 3500 rpm for 5 minutes after clotting. The serum was separated for testing within 3 hours using a LumipulseG1200 (Japan) system fully automated chemiluminescent immunoassay analyzer and its supporting reagents. The method was a two-step sandwich chemiluminescence enzyme immunoassay, which detected the maximum luminescence intensity to reflect the amount of PIVKA-II in the coupled particles at a wavelength of 477 nm. The manufacturer provided the LumiPulse G PIVKA-II calibrators (FUJIREBIO INC. Tokyo 163-0410, Japan). The calibration solution was traceable to the company's

internal standards. It could be traced back to the PIVKA-II ECLIA kit (Picolumi PIVKA-II, EIDIA Co. Ltd) with a good correlation coefficient (r=0.992, y=0.98x-143.60, n=152). At least two levels of quality control (LumiPulse PIVKA-II Quality Control) were tested with the samples every working day. The instrument's detection range was 5~75000 mAU/ml, with intra-batch precision ≤ 3.5% and inter-batch precision ≤ 4.5%.

Storage group data screening

The raw data included all testing values of the storage group. The authors obtained the final valid data for statistical analysis by screening with the following rules. (1) Screen out multiple test values of the same patient, excluding duplicate data and retaining only the initial test results. (2) Screen out data from tumour, hepatobiliary, and infection departments. (3) Screen out data of death cases and data of confirmed liver-related diseases. (4) Screen out data of postoperative tumour, suspected tumour, and diagnosed tumour. (5) Screen out data of which source of hypoproteinemia could not be determined. (6) The quartile values of testing values P25 (Q1) and P75 (Q3) were calculated to obtain the quartile spacing IQR (Q3 - Q1), and outliers were removed, with the criterion that those more excellent than Q3 + 3IQR excluded.

Calculating the indirect reference intervals of the storage group data

The author partitioned the storage group data into gender-specific subgroups, and four methods were used to establish the indirect reference intervals for the subgroups.

Firstly, the author used piece-wise regression to determine the linear portion of the cumulative frequency graph, named the Hoffmann method (HM)[11,12]. The idea behind piecewise linear regression was that if the data followed different linear trends over different regions of the data, then we should model the regression function in "pieces". Linear regression could be denoted by an equation ($f(c_i) = \alpha * c_i + \beta$) for the linear portion in each partition (α was the slope, β was the intercept of the line, c_i was the cumulative frequency). The breakpoint was determined by the minimum value of the sum of squared residuals through continuous iterations. The sum of squared residuals was defined as:

$$\mathrm{SSE} = \sum(\mathrm{Y_i} - \mathrm{f(c_i)})^2, \qquad (1)$$

where Y_i was the observed value and $f(c_i)$ was the linear regression function.

The authors also determined the best-fitting piece-wise regression by least-squares analysis. The right-sided reference upper limit was calculated by extrapolating the 95th percentiles in the linear portion. Finally, the authors drew the cumulative frequency chart and curve in Figure 1 and the piece-wise regression curve in Figure 2.

In Figure 1, the authors found that the storage group data had a skewed distribution, which might not meet the conditions for the Hoffmann method. A revised version of the Hoffmann method (HMCDF) was carried out to solve this problem, which determined the linear portion with inverse CDF piece-wise regression. The authors calculated the inverse cumulative distribution function (CDF) of a standard Gaussian distribution for each PIVK-II value[13]. The testing values were plotted against the

199 inverse CDF of the standard Gaussian distribution in Figure 2. Similarly, the authors
200 used piece-wise regression to determine the linear portion of the resulting graph. The
201 breakpoint and best-fitting regression functions were determined by minimizing the
202 sum of squared residuals. The right-sided reference upper limit was calculated by
203 extrapolating the 95th percentiles based on the inverse CDF values corresponding to
204 1.645 of the standard Gaussian distribution.
205
206 Thirdly, the authors chose a statistical method (EMBCT)[14] to establish the reference
207 interval, which assumed that the storage group data was sampled from two
208 distributions, the healthy and diseased distribution. The basic idea of the EMBCT was
209 to separate these two distributions. The observed PIVK-II testing data Y_i can be
210 derived as:
211 $$Y_i = U_i X_i^1 + (1 - U_i) X_i^2, i = 1,2, \dots, n, \qquad (2)$$
212 where the authors used U_i to denote the status of the ith individual equal to 1 when
213 the individual was healthy and 0 otherwise, X_i^1 was the testing data when the
214 individual was healthy and X_i^2 when the individual was diseased. The model
215 assumed that the random variables X_i^j in equation (2) were mutually independent,
216 independent of U_i and after a Box-Cox transformation, X_i^1 and X_i^2 obeyed two
217 different normal distributions. The Box-Cox transformation was defined as:
218 $$k_\lambda(x) = \begin{cases} \frac{x^\lambda - 1}{\lambda} & if \quad \lambda \neq 0 \\ \log(x) & if \quad \lambda = 0, \end{cases} \qquad (3)$$
219 where λ was a parameter. In practice, U_i cannot be observed; instead, one can use a
220 parameter p to denote the proportion distribution of the healthy and diseased

distributions. The likelihood function can be derived as:

$$L(\theta) = \prod_{i=1}^{n}(p\gamma_1(Y_i) + (1-p)\gamma_2(Y_i)), \qquad (4)$$

where $\gamma_1(Y_i)$ denoted the distribution of Y_i when the individual was healthy, $\gamma_2(Y_i)$ denoted the distribution of Y_i when the individual was diseased, and θ represented all the parameters including p and parameters in $\gamma_1(Y_i)$ and $\gamma_2(Y_i)$. To derive the distribution of PIVK-II in healthy individuals, the authors calculated the maximum likelihood estimate of θ using the EM algorithm[15] and BFGS algorithm[16]. The distribution of PIVK-II in healthy individuals can be denoted by:

$$\gamma_1(Y_i) = \frac{{Y_i}^{\lambda_1-1}}{\sigma\sqrt{2\pi}} exp\left[-\frac{1}{2\sigma^2}(k_{\lambda_1}(Y_i) - m)^2\right], \qquad (5)$$

where m and σ^2 are expectation and variance of $k_\lambda(Y_i)$ and λ_1 is the parameter in Box-Cox transformation. One can derive the right-sided reference interval as:

$$\left[0,\ (1+\hat{\lambda}_1(\hat{m}+1.645\hat{\sigma}))^{\frac{1}{\hat{\lambda}_1}}\right], \qquad (6)$$

where $\hat{m}$, $\hat{\sigma}$ and $\hat{\lambda}_1$ are estimations of m, σ and λ_1. The author drew the $\gamma_1(Y_i)$ and $\gamma_2(Y_i)$ In Figure 3, and all the programs were coded by python.

Fourthly, the authors adopted a recent method named KOSMIC[17] to construct the reference intervals. The method's parameters assumed that the proportion of healthy samples in the input dataset could be modelled with a Gaussian distribution after the Box-Cox transformation of the data. A truncation interval T existed within the dataset, in which the proportion of diseased test results is negligible.

The algorithm aimed to minimize the Kolmogorov-Smirnov distance between an estimated normal distribution F, and a truncated part of the observed distribution of

test results after Box-Cox-transformation D. The Kolmogorov-Smirnov distance is denoted by:

$$KS = \frac{\sup|D-F|}{\sqrt{n}} + P_1 + P_2, \qquad (8)$$

where D denotes the cumulative density function of the dataset after Box-Cox transformation using λ, F represents the cumulative density function of a normal distribution described by μ and σ, and n denotes the number of samples within T. P_1 and P_2 are penalty terms for test results outside the truncation interval, which are defined as:

$$P_1 = \frac{\sup F-D}{\sqrt{n}}, \qquad (9)$$

$$P_2 = \frac{\sup D-F}{\sqrt{n}}. \qquad (10)$$

The parameters of the normal distribution (μ, σ), the Box-Cox-transformation parameter (λ), and the truncation interval T are optimized numerically, consistent with previous work[17]. To provide confidence intervals, the authors used bootstrapping of the original dataset (random sampling with replacement). 90% PIVKA-II reference upper limit confidence intervals for the storage group calculated by HM, HMCDF and KOSMIC were presented.

Finally, we used the model averaging technique to estimate the reference upper limit for PIVKA-Ⅱ. Model averaging refers to the process of evaluating some quantity under each model and then averaging the estimates according to how likely each model is. We estimated the reference upper limit $\hat{Y}$ using model averaging with the following equation:

$$\hat{Y} = \sum_{i=1}^{s} \hat{Y}_i \Pr(M_i|y^n), \qquad (11)$$

where M_i denoted HM, HMCDF, EMBCT and KOSMIC, y^n was the storage group data for PIVKA-Ⅱ, $\hat{Y}_i$ was the estimator of M_i. $\Pr(M_i|y^n)$ could be approximated by[18,19]:

$$\Pr(M_j|y^n) = \frac{e^{\widehat{m}_j}}{\sum_{i=1}^{s} e^{\widehat{m}_i}}, \qquad (12)$$

$$\widehat{m}_i = \hat{l}_i - \frac{d_i}{2}\log n, \qquad (13)$$

where $\hat{l}_i$ is the log-likelihood function for model M_i and d_i denoted the dimension of the parameters in the model M_i.

For HM and HMCDF, piece-wise regression was used to estimate the reference upper limit. We have PIVKA-Ⅱ data on n subjects and a model of the form:

$$y_i = \begin{cases} \beta_0 + \beta_1 x_i + \varepsilon_i & x_i > c \\ \beta_2 + \beta_3 x_i + \varepsilon_i & x_i \le c \end{cases}, \qquad (14)$$

Where ε_i obeys $N(0, \sigma^2)$ independently. For each linear portion, we could calculate the log-likelihood:

$$\hat{l} = -\frac{n}{2}\log(2\pi) - n\log\sigma - \sum_{i=1}^{n} \frac{1}{2\sigma^2}(y_i - \beta_0 - \beta_1 x_i)^2. \qquad (15)$$

Inserting the MLEs of β and σ into the above equation, we found that $\widehat{m}_i$ became:

$$\widehat{m}_i = -\frac{n}{2}\log(2\pi) - n\log\hat{\sigma}_s - \frac{n}{2} - \frac{d_i}{2}\log n, \qquad (16)$$

where $\hat{\sigma}_s$ was the MLE of σ. For HM, we could calculate $\widehat{m}_i$ using equation (16). In HMCDF, x_i represented the PIVKA-Ⅱ value and y_i was the inverse CDF of a standard Gaussian distribution. To calculate $\widehat{m}_i$ for HMCDF properly, we shifted x_i and y_i, and recalculated the parameters of the piece-wise regression. For EMBCT, the log-likelihood could be derived as:

$$\hat{l} = \sum_{i=1}^{n} \log(p\gamma_1(y_i) + (1-p)\gamma_2(y_i)). \quad (17)$$

$\gamma_1(y_i)$ and $\gamma_1(y_i)$ were consistent with equation (5). For KOSMIC, the log-likelihood was denoted by:

$$\hat{l} = \sum_{y_i \ within \ T} \log(\gamma(y_i)). \quad (18)$$

After calculating the $\hat{m}_i$ value, $\Pr(M_i|y^n)$ was approximated with equation (12) and the reference upper limit could be estimated with equation (11).

Statistical analysis of health group data

According to the non-parametric method recommended in CLSI document C28-A3, (1) we removed outliers (Tukey's method), with the criterion that those greater than Q3 + 3IQR excluded. (2) As reference intervals might be affected by age, gender and ethnicity, we partitioned the health group data into different subgroups according to gender and age. (3) Harris and Boyd's test was used to evaluating the difference in reference intervals between the subgroups. These subgroups can be merged if no significant difference is detected between the adjacent age-specific orgender-specific subgroups. (4) We established the 95th percentiles as the reference upper limit and constructed the 95% right-sided reference intervals for each subgroup. In addition, 90% PIVKA-II reference upper limit confidence interval for the health group was also calculated.

Results

PIVKA-II's original data contained 7623 cases, and the valid data were 1152 cases

after three screenings, with a ratio of 15.11%. Details could be found in Table 1.

Establishment of storage group reference intervals

The authors drew the cumulative frequency distribution chart and curve of gender-specific subgroups for the storage group data in Figure 1. The piece-wise regression curves for the gender-specific subgroups calculated by HM and HMCDF were drawn in Figure 2. In Figure 3, the authors plotted the cumulative frequency graph and the density function curves of $\gamma_1(Y_i)$ and $\gamma_2(Y_i)$ for the gender-specific subgroups. PIVKA-II's indirect reference intervals calculated with four different methods in the storage data group were respectively 0~35.30 mAU/mL, 0~31.48 mAU/mL, 0~30.78 mAU/mL, and 0 to 36.17 mAu/mL. The averaging result over these four models was 0~31.48 mAU/mL. Compare with HM, EMBCT and KOSMIC, HMCDF was the most probably model. Table 2, Table 3 and Table 4 detail the indirect reference intervals of PIVKA-II based on the storage group data. In Table 5, the authors presented 90% PIVKA-II reference upper limit confidence intervals for the storage group calculated by HM, HMCDF and KOSMIC.

Establishment of reference intervals for the healthy group

No outliers were included in the healthy group, with the criterion that values greater than Q3 + 3IQR excluded (Tukey method). The authors found no significant difference between the adjacent age-specific male subgroups, and all the age-specific male subgroups could be merged. However, PIVKA-II showed a distinction between

the adjacent age-specific female subgroups, indicating that the age-specific female subgroups could not be combined. Furthermore, a significant difference was detected between the gender-specific subgroups, which suggested that gender impacted the PIVKA-II value. The reference intervals calculated from storage and health groups data showed that the female group's reference upper limit was smaller than the male group.

The authors drew the cumulative frequency distribution chart and curve of PIVKA-II for the healthy group and gender-specific subgroups in Figure 1. PIVKA-II's 95% right-sided reference interval was 0~32.00 mAU/mL, and the 90% PIVKA-II reference upper limit confidence interval was 31.00~32.60 mAU/mL. Age-specific reference intervals, gender-specific reference intervals, the median and IQR (interquartile range) were calculated in Table 6.

Validation of reference intervals for the healthy group

The authors randomly chose 20 healthy individuals with physical examinations in the laboratory from January to February 2018 as study subjects. All specimens met the corresponding conditions above. One out of the nineteen cases' serum PIKVA-II concentration exceeded the indirect reference intervals established by the indirect methods, which gives the reference interval a 95% specificity demonstrating its validity. The reference intervals constructed by the indirect method met the laboratory quality management requirements and can be used for clinical diagnosis.

Discussion

The reference interval of a biomarker is one of the dimensions for defining disease status, which plays an essential role in clinical diagnosis and treatment. However, due to some limitations, many clinical laboratories choose to use manufacturers' reagent instructions as reference intervals. Ignoring that reference intervals may vary with region, ethnicity, and age can result in the problem that the adopted reference intervals are not accurate enough to fulfil its clinical needs. A recent development in LIS in hospitals makes it possible for researchers to access and cross-reference data to explore its value. One of the ideas is to utilize the LIS storage data to establish biological reference intervals. Over the past decades, this idea has been developed and improved, and a relatively mature methodological system has been established [20-22]. As early as 1963, some researchers established serum glucose reference intervals by Hoffmann method[11], which screened hospital mixed data and made an initial attempt to establish reference intervals with hospital storage data. Then the value of hospital storage data was gradually discovered.

This study used four methods to calculate PIVKA-II's indirect reference intervals from LIS's stored data.

The first method was the Hoffmann method, which determined the linear portion with piece-wise regression. The indirect reference interval of PIVAK-II irrespective of gender and age was 0-35.30, and the indirect reference interval of PIVAK-II according to different genders was 0-35.97 for males, 0-32.29 for females. The Hoffmann method required that the hospital data for testing formed a Gaussian

distribution, and most of the testing made in the hospital represented normal individuals. However, in Figure 1, the author found that the storage group data collected had a skewed distribution. Therefore, the authors revised the Hoffmann method, which determined the linear portion with inverse CDF piece-wise regression. The indirect reference interval of PIVAK-II irrespective of gender and age was 0-31.48, and the indirect reference interval of PIVAK-II according to different genders was 0-31.97 for males, 0-30.96 for females. The third method was EM algorithm-based method, which can be considered as a gold-standard method for mixtures analysis[14]. After a Box-Cox transformation, it did not require that the hospital-based data obeyed Gaussian mixture distribution. As a result, the indirect reference interval of PIVAK-II irrespective of gender and age was 0-30.78, and the indirect reference interval of PIVAK-II according to different genders was 0-30.92 for males, 0-30.65 for females. The fourth method was a recent estimator using truncation points and the Kolmogorov-Smirnov distance (KOSMIC), which assumed that the proportion of healthy samples in the input dataset could be modelled with a Gaussian distribution after Box-Cox transformation of the data. The indirect reference interval of PIVAK-II irrespective of gender and age was 0-36.17, and the indirect reference interval of PIVAK-II according to different genders was 0-37.40 for males, 0-32.35 for females.

The Hoffmann method did not perform well when handling data that obeys a skewed distribution because of the Gaussian distribution requirement. KOSMIC assumed that a truncation interval T existed within the dataset, in which the proportion of diseased

test results is negligible. However, the truncation points assumption might not correspond to the actual situation. HMCDF and EMBCT did not have the Gaussian distribution requirement or truncation points assumption. After averaging the estimators over four models, we found that HMCDF was the most probably model and the averaging result was the same as the HMCDF result. Furthermore, comparing with the other indirect methods, the reference interval established by HMCDF and model averaging was most close to the result obtained from the health group data.

PIVKA-II's reference interval obtained from the health group data was 0~32.00 mAU/mL, which could be considered an accepted classic method based on the CLIS C28-A3 guidelines. There was no evident difference between the reference interval obtained from the health group and the indirect reference intervals established by the model averaging. Furthermore, Table 2, Table 3, and Table 4 presented that the female group's reference upper limit was smaller than the male group. Published research showed that the incidence of HCC was significantly higher in males than females[23], and PIKVA-II was elevated in a certain percentage of HCC patients[6]. Compared with the male group, the minor reference upper limit we obtained for the female group was consistent with previous studies. The reference intervals calculated by HMCDF for the gender-specific subgroups were also close to the health group data results. After comparing different approaches, HMCDF and model averaging techniques were recommended to establish the indirect reference intervals for PIKVA-II. Data screening and gender-specific subgroups needed to be considered during the process.

The 90% confidence interval of the reference upper limit established based on CLSI

C28-A3 (31.00~32.60 mAU/mL) did not cover the reference upper limit of the manufacturer's reagent instructions (0~40 mAU/mL), which suggested these two reference intervals were different. The difference might be caused by the fact that the manufacturer's reagent instructions mainly were based on European and North American populations, and the ethnic and regional differences led to the interval difference. However, the 90% confidence interval of the reference upper limit calculated by HMCDF covered the reference upper limit established based on CLSI C28-A3, suggesting no significant difference between the storage group's indirect reference interval the health group data results. Therefore, the indirect reference interval established in this study could be used to validate the existing clinical reference interval to some extent.

The application of storage data to establish biological reference intervals for biomarkers has outstanding advantages. To begin with, it has the advantage of being compatible with the population corresponding to the reference interval. The LIS's storage data has characteristics corresponding to the local populations, including ethnicity, geography, and food habits. Furthermore, it can also save a lot of labour and material resources to improve efficiency. CLSI document C28-A3 indicates that the study individuals should be healthy, but the clinical patients bedridden for a long time significantly differ from healthy individuals. Therefore, collecting data from a healthy population is costly and time-consuming. However, indirect methods make full use of the LIS's stored data and provide a new direction for establishing laboratory reference intervals.

Reference

1. Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018:GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries [J]. CA: A Cancer Journal for Clinicians, 2018, 68(6): 394-424.
2. Wang X, Zhang W, Liu Y, et al. Diagnostic value of prothrombin induced by the absence of vitamin K or antagonist-II (PIVKA-II) for early-stage HBV related hepatocellular carcinoma[J]. Infectious Agents and Cancer, 2017, 12(1): 1-8.
3. Lok AS, Sterling R K, Everhart J E, et al. Des-gamma-carboxy prothrombin and alpha-fetoprotein as biomarkers for the early detection of hepatocellular carcinoma [J]. Gastroenterology, 2010, 138(2): 493-502.
4. Park S J, Jang JY, Jeong S W, et al. Usefulness of AFP, AFP-L3, and PIVKA-II, and their combinations in diagnosing hepatocellular carcinoma[J]. Medicine, 2017, 96(11).
5. Yu R, Tan Z, Xiang X, et al. Effectiveness of PIVKA-II in detecting hepatocellular carcinoma based on real-world clinical data[J]. BMC Cancer, 2017, 17(1): 1-10.
6. Liebman H A, Furie BC, Tong M J, et al. Des-γ-carboxy (abnormal) prothrombin as a serum marker of primary hepatocellular carcinoma[J]. New England Journal of Medicine, 1984, 310(22): 1427-1431.
7. Tsuchiya N, Sawada Y, Endo I, et al. Biomarkers for the early diagnosis of hepatocellular carcinoma[J]. World Journal of Gastroenterology: WJG, 2015,

21(37): 10573.

8. Horowitz G, Altaie S, Boyd J, et al. Clinical and Laboratory Standards Institute: Defining, Establishing, and Verifying Reference Intervals in the Clinical Laboratory; Approved Guideline[J]. Wayne, PA: CLSI, 2008.
9. Ozarka, Yesim.Reference intervals: current status, recent developments and future considerations. Biochemia Medica,2016,26(1):5-16.
10. Arzideh, F, Wosniok, W, Haeckel, R. Indirect reference intervals of plasma and serum thyrotropin (TSH) concentrations from intra-laboratory databases from several German and Italian medical centres.Clinical Chemistry and Laboratory Medicine.2011,49(4):659-664.
11. Katayev A, Balciza C, Seccombe D W. Establishing reference intervals for clinical laboratory test results: is there a better way? [J]. American journal of clinical pathology, 2010, 133(2): 180-186.
12. Hoffmann R G. Statistics in the practice of medicine[J]. Jama, 1963, 185(11): 864-873.
13. Shaw J L V, Cohen A, Konforte D, et al. validity of establishing pediatric reference intervals based on hospital patient data: a comparison of the modified Hoffmann approach to CALIPER reference intervals obtained in healthy children[J]. Clinical Biochemistry, 2014, 47(3): 166-172.
14. Concordet D, Geffré A, Braun J P, et al. A new approach for the determination of reference intervals from hospital-based data[J]. Clinica Chimica Acta, 2009, 405(1-2): 43-48.

15. Dempster AP, Laird NM, Rubin D B. Maximum likelihood from incomplete data via the EM algorithm[J]. Journal of the Royal Statistical Society: Series B (Methodological), 1977, 39(1): 1-22.
16. Liu DC, Nocedal J. On the limited memory BFGS method for large scale optimization[J]. Mathematical Programming, 1989, 45(1): 503-528.
17. J Zierk, Arzideh F, Kapsner L A, et al. Reference Interval Estimation from Mixed Distributions using Truncation Points and the Kolmogorov-Smirnov Distance (Kosmic)[J]. Scientific Reports.
18. Kass RE, Wasserman L. A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion[J]. Journal of the American Statistical Association, 1995, 90(431): 928-934.
19. Kass RE, Raftery A E. Bayes factors[J]. Journal of the American Statistical Association, 1995, 90(430): 773-795.
20. Arzideh F, Brandhorst G, Gurr E, et al. An improved indirect approach for determining reference limits from intra-laboratory data bases exemplified by concentrations of electrolytes/Ein verbesserter indirekter Ansatz zur Bestimmung von Referenzgrenzen mittels intra-laboratorieller Datensätze am Beispiel von Elektrolyt-Konzentrationen[J]. Journal of Laboratory Medicine, 2009, 33(2): 52-66.
21. Zierk J, Arzideh F, Rechenauer T, et al. Age-and gender-specific dynamics in 22 hematologic and biochemical analytes from birth to adolescence[J]. Clinical Chemistry, 2015, 61(7): 964-973.

22. Jones G R D, Haeckel R, Loh TP, et al. Indirect methods for reference interval determination–review and recommendations[J]. Clinical Chemistry and Laboratory Medicine (CCLM), 2018, 57(1): 20-29.
23. El-Serag H B. Hepatocellular carcinoma: recent trends in the United States[J]. Gastroenterology, 2004, 127(5): S27-S34.

Figure 1: (A) cumulative frequency graph for the storage group. (B) cumulative frequency graph for the male storage subgroup. (C) cumulative frequency graph for the female storage subgroup. (D) cumulative frequency graph for the health group. (E) cumulative frequency graph for the male health subgroup. (F) cumulative frequency graph for the female health subgroup.

Figure 2: (A) Cumulative frequencies (dots) and the piece-wise regression curve for the storage group data. (B) Cumulative frequencies (dots) and the piece-wise regression curve for the male subgroup. (C) Cumulative frequencies (dots) and the piece-wise regression curve for the female subset. (D) inverse CDF of a standard Gaussian distribution against each testing value and the piece-wise regression curve for the storage group data. (E) inverse CDF of a standard Gaussian distribution against each testing value and the male subgroup's piece-wise regression curve. (F) inverse CDF of a standard Gaussian distribution against each testing value and the female subgroup's piece-wise regression curve.

Figure 3: (A) cumulative frequency graph and the density function curves of $\gamma_1(Y_i)$ and $\gamma_2(Y_i)$ for the storage group data. (B) cumulative frequency graph and the density function curves of $\gamma_1(Y_i)$ and $\gamma_2(Y_i)$ for the male subgroup. (C) cumulative frequency graph and the density function curves of $\gamma_1(Y_i)$ and $\gamma_2(Y_i)$ for the female subgroup.

Figures

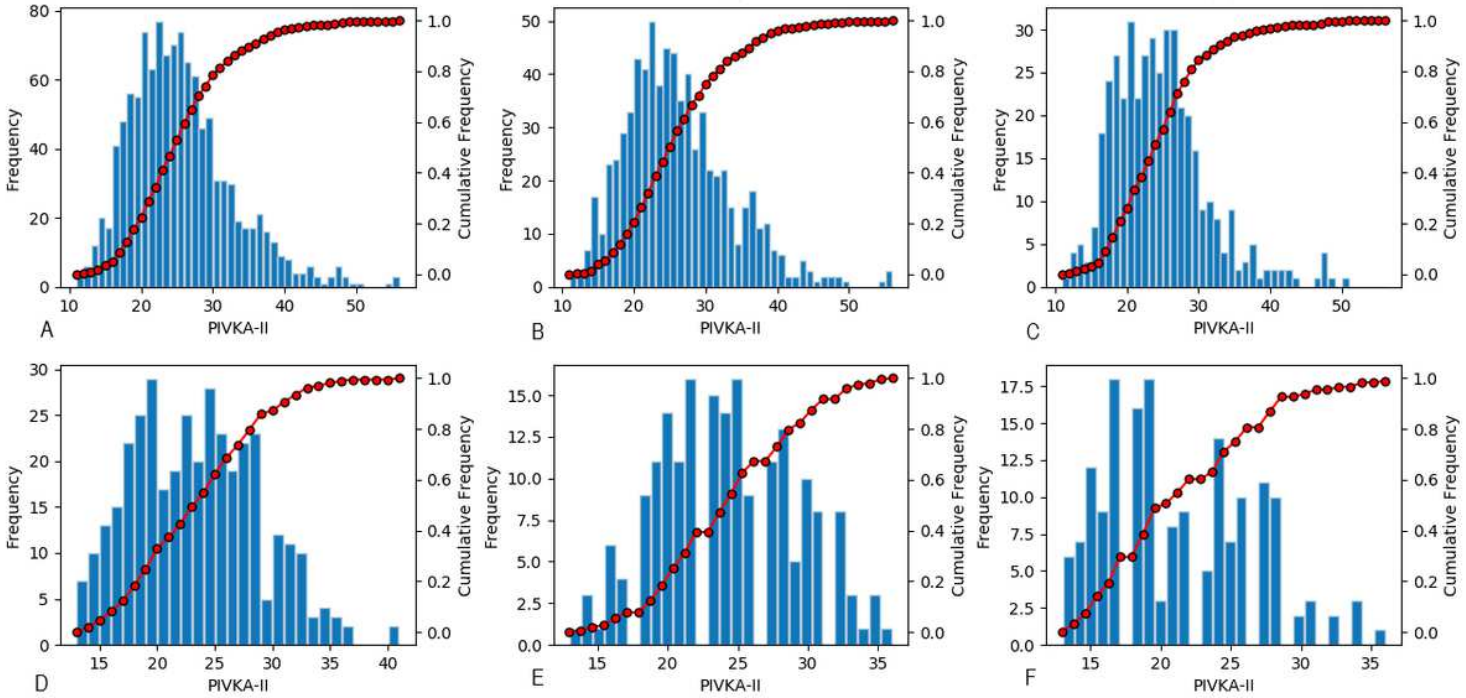

Figure 1

(A) cumulative frequency graph for the storage group. (B) cumulative frequency graph for the male storage subgroup. (C) cumulative frequency graph for the female storage subgroup. (D) cumulative frequency graph for the health group. (E) cumulative frequency graph for the male health subgroup. (F) cumulative frequency graph for the female health subgroup.

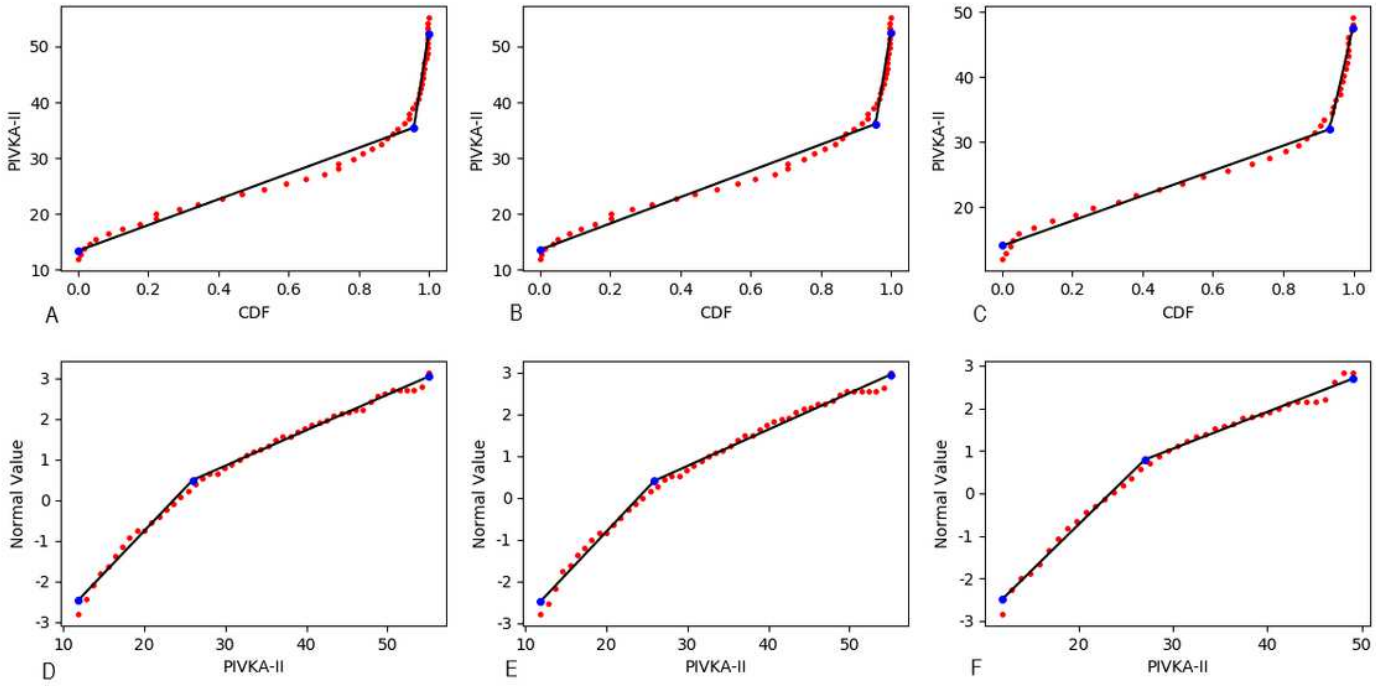

Figure 2

(A) Cumulative frequencies (dots) and the piece-wise regression curve for the storage group data. (B) Cumulative frequencies (dots) and the piece-wise regression curve for the male subgroup. (C) Cumulative frequencies (dots) and the piece-wise regression curve for the female subset. (D) inverse CDF of a standard Gaussian distribution against each testing value and the piece-wise regression curve for the storage group data. (E) inverse CDF of a standard Gaussian distribution against each testing value and the male subgroup's piece-wise regression curve. (F) inverse CDF of a standard Gaussian distribution against each testing value and the female subgroup's piece-wise regression curve.

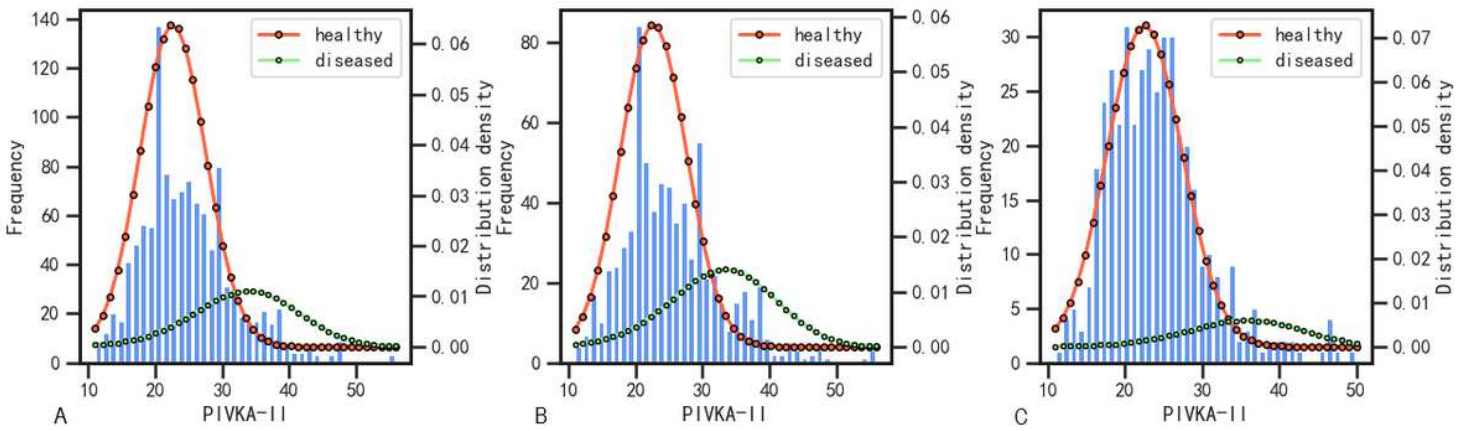

Figure 3

(A) cumulative frequency graph and the density function curves of γ_1 (Y_i) and γ_2 (Y_i) for the storage group data. (B) cumulative frequency graph and the density function curves of γ_1 (Y_i) and γ_2 (Y_i) for the male subgroup. (C) cumulative frequency graph and the density function curves of γ_1 (Y_i) and γ_2 (Y_i) for the female subgroup.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- table3.docx
- table4.docx