Iran J Public Health, Vol. 53, No.5, May 2024, pp.1192-1199

Original Article

A New Formula Based on Simple Blood Indices to Differentiate Beta Thalassemia Trait from Iron Deficiency Anemia

*Parviz Shahmirzalou [1], Morteza Seyyed Hamze [2], *Hassan Ebrahimpour Sadagheyani [3]*

1. Department of Biostatistics, Khoy University of Medical Sciences, Khoy, Iran
2. Urmia Branch, Islamic Azad University, Urmia, Iran
3. Department of Health Information Technology, School of Para Medicine, Neyshabur University of Medical Sciences, Neyshabur, Iran

***Corresponding Authors:** Emails: sadageyani@yahoo.com, ebrahimpourh@nums.ac.ir

(Received 21 May 2023; accepted 26 Aug 2023)

Abstract

Background: Thalassemia is an inherited blood disorder with a defect in the sufficient production of a protein called hemoglobin. We aimed to investigate the simple blood indices of patients with Beta Thalassemia Trait (BTT) and Iron Deficiency Anemia (IDA) to propose a new formula using logistic regression for differentiate two characteristics from each other.

Methods: Among the 702 records of the BTT Counseling Center (Khoy, Iran-2022), 292 cases (219 iron deficiency anemia (IDA) and 73 BTT) were eligible for the study. Blood indices such as RBC, HGB, HbA2 described and used to diagnose two types of participants. Blood indices had high multicollinearity that was modified. Logistic regression for blood indices fitted and goodness of fit indices with Area Under ROC curve (AUC) estimated.

Results: The average age of the participants was 24.56 yr. The status of Multicollinearity between independent variables was modified. The HGB, MCV, HbA2, and HbA variables were used in the model and only HbA2 status was significant ($P<0.001$). According to the output of the model, for each unit increase in HbA2, the chance of having BTT was about 8.5 times higher than IDA. The sensitivity, specificity, AUC curve, and accuracy of the final model were estimated to be 97, 72, 84, and 93%, respectively. A regression formula to differentiate BTT from IDA proposed.

Conclusion: In studies related to the differentiation of the BTT from IDA, the presence of the HbA2 index in the model and prediction is very necessary.

Keywords: Beta thalassemia; Iron deficiency anemia; Logistic regression; Blood index; Differentiation formula

Introduction

Thalassemia is an inherited blood disorder with a defect in the sufficient production of a protein called Hemoglobin (HGB). Under insufficiency of HGB, the function of Red Blood Cells (RBCs) is impaired and their lifespan is shortened. RBCs were necessary to transfer oxygen to cells. Then if RBCs die early, the patient has not enough oxygen and feels tired with weakness. His breathing becomes short, and we say that the person was suffering from thalassemia (or anemia). Thalassemia can be mild to severe, the severe type of which is traumatic. The type of thalassemia depends on the damaged part of hemoglobin (alpha or beta) and the severity of damage (minor, intermediate, major). Clinical symptoms are mild in thalassemia minor and severe in thalassemia major. In a study by Centers for Disease Control and Prevention, the role of demographic predictors and blood indices in the development of beta-thalassemia minor is investigated (1).

According to the WHO report, thalassemia affects about 46 cases per 100,000 people in the world and about 3.4% of births. Of 330,000 babies born with hemoglobin disorders, nearly 17% are affected with thalassemia. Moreover, 7% of pregnant women carry hemoglobin disorders (2). The alpha and beta forms of thalassemia are more common than other forms, and beta thalassemia is very common in Mediterranean countries, Africa and Central Asia. According to the report (3), about 0%-3% of the America, 2%-18% of the Eastern Mediterranean, about 0%-19% of Europe, 0%-11% of Southeast Asia, 0%-12% of sub-Saharan Africa, and 0%-13% of the Western Pacific have Beta Thalassemia (BT) gene mutation. In Asian countries, the situation of BT varies from one country to another. Khodayi et al. reported the prevalence of BT in the Khorasan-Razavi Province of Iran as 0.6% (4). Miri et al. (5) presented a comprehensive article on the prevalence of beta-thalassemia and the success of the beta-thalassemia prevention program from 1996 to 2009. They reported that during the years 1989-2009, about 12750 babies with thalassemia were born. The prevention percentage of thalassemia in 1996 was about 26.3% and in 2009, it reached 82.3%. In 2009, about 239 babies with thalassemia were born and the birth of 745 ill babies was prevented. In this study, the average prevalence rate of thalassemia in Iran has been reported as 4%. Kerman and West Azerbaijan provinces have the highest and lowest statistics with a prevalence of 9.5% and 1%, respectively (5).

So far, various studies have been conducted to identify couples and babies with BT problems (6-10) and several formulas have been introduced for the differential diagnosis of beta-thalassemia (11, 12). At the time of marriage, couples refer to the BT prevention centers in the city and undergo the necessary tests and consultations to determine their condition in terms of minor BT. This happens before (a few days before) the marriage, while in Iranian culture, between courtship and referral to the beta-thalassemia counseling center, various meetings are held between the couple's families, and some decisions about the future of the couples were done. Therefore, it becomes impossible or difficult for families to break up the marriage process due to beta-thalassemia. Now the question is whether it is possible to find out whether a person has problems with beta-thalassemia based on the basic blood indicators (RBC, HCT, etc.) that are reported in most routine tests. According to aforementioned details, a study was conducted on the couples referred to the BT-counseling center.

We aimed to investigate the blood indices of BT carriers and healthy couples referring to this center to propose a formula with logistic regression for simple blood indices and differentiate Beta Thalassemia Trait (BTT) from Iron Deficiency Anemia (IDA).

Materials and Methods

The data of this study are related to the medical records of Khoy beta-thalassemia counseling center, Iran. For this purpose, BT files of all patients (702 cases) were examined at the beta-thalassemia counseling center (Khoy, Iran-2022). If the person had hypochromic-microcytic anemia, he/she would be included in the study (inclusion criteria). In hematology, a person with MCV less than 80fl and MCH less than 27pg is hypochromic-microcytic anemia (6). According to the national BT guidelines, couples with hypochromic-microcytic anemia referred to the provincial reference center for additional tests to definitively determine whether they are carriers of BT. Based on additional tests, 292 patients were eligible to enter the study, among them, 219 hypochromic-microcytic anemia and 73 BT carriers were diag-

nosed. Since patients have two statuses, we intended to use logistic regression for analysis. Logistic regression was classified in the set of Generalized Linear Models (GzLM) to investigate the relationship between the dependent variable (BTT or IDA) with blood indices, plot the ROC curve and evaluate the ability of the model to differentiate BT from hypochromic-microcytic anemia with the measure of Area Under the ROC Curve (AUC). In fitting the model, blood indices such as Red Blood Cell (RBC), Hemoglobin (HGB), Hematocrit (HCT), MCV, MCH, Red cell Distribution Width (RDW), Hemoglobin A2 (HbA2), Hemoglobin A(HbA) and Ferritin were used. The goodness of fitted logistic regression was checked with three fitting indices. The first criterion is the omnibus test of model coefficients that surveys the significance of regression coefficients for all independent variables. Rejecting the null hypothesis was desirable. The second fit criterion was the Nagelkerke R-square index, which is similar to the R-square index in linear regression. The larger its value, the greater the ability of the predictor variables to express variations in the dependent variable. Finally, the third goodness-of-fit criterion was the Hosmer-Lemeshow test. It measures the correct classification ability of the model (classification in the study and the one predicted by the model).
Accepting the null hypothesis in this test indicated the ability of the model to optimally classify patients into dependent variable subgroups. Another potency of logistic regression was Odds Ratio (OR). The interpretation of regression coefficients in logistic regression is done by OR index. Multicollinearity was one of the multiple regression assumptions. The amount of multicollinearity between independent variables be less than 10 or mean of multicollinearity was less than one (13).
The protocol of study was approved by the Ethical Committee of the Research Deputy of Khoy University of Medical Sciences, Khoy, Iran (IR.KHOY.REC.1398.004). Informed consent had been taken from all patients.

Results

There were 292 participants in the study. The average (standard deviation) age of the participants was 24.56 (6.02) yr. About half of the participants were under 24 yr old. The number of men was equal to the number of women. Fig. 1 reports the frequency of participants education adjusted to sex of them.

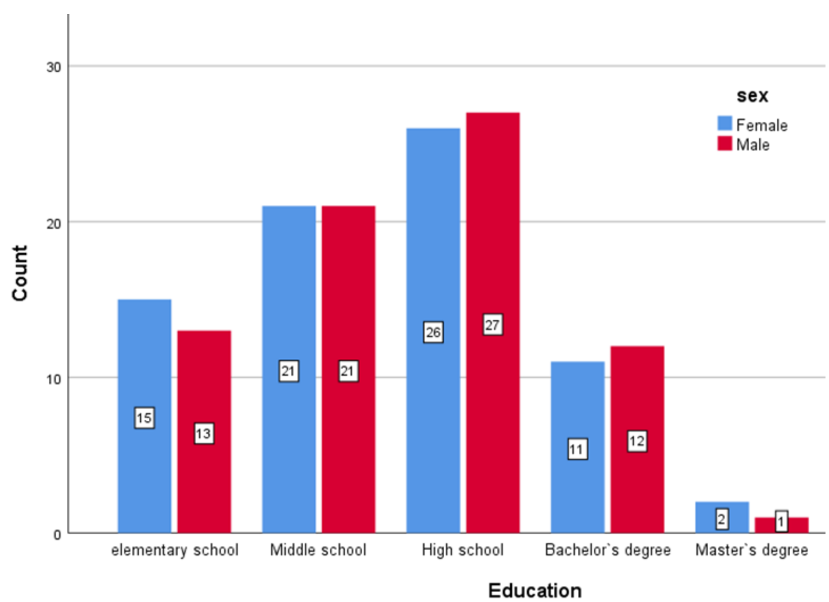

Fig. 1: Bar plots of participants education in the study

Descriptive statistics and equality of mean test for blood indices in IDA and BTT groups are reported in Table 1. The number of missing cases was high for some blood indices (HbA, Ferritin), which were retained in the data analysis due to the high importance of the index.

Table 1: Description of blood indices and test about equality of mean for IDA and BTT groups

Index (N valid)	*Total*		*IDA*	*BTT*	*P-value*
	Mean (SD)	Min (Max)	Mean (SD)	Mean (SD)	
RBC (288)	5.64 (0.72)	3.98 (7.86)	5.48(0.63)	6.13(0.77)	<0.001*
HGB (286)	13.45 (1.92)	8.5 (17.90)	13.74(1.97)	12.51(1.37)	<0.001*
HCT (286)	42.24 (4.58)	29.50 (52.90)	42.48(4.67)	41.46(4.20)	0.112*
MCV (290)	75.12 (7.11)	39.60 (87.90)	77.52(5.69)	67.83(5.93)	<0.001**
MCH (290)	24.00 (3.20)	16.90 (29.70)	25.13(2.57)	20.57(2.37)	<0.001**
RDW (283)	14.70 (2.47)	12 (38.80)	14.37(2.59)	15.75(1.63)	<0.001**
HbA2 (292)	3.17 (1.39)	0.20 (7.40)	2.54(0.78)	5.06(1.07)	<0.001**
Ferritin (140)	54.35 (57.93)	1.40 (333.30)	51.05(55.01)	93.11(78.22)	0.036**
HbA (174)	94.39 (12.35)	2.06 (98.31)	95.06(11.05)	90.86(17.51)	<0.001**

Nominal logistic regression was fitted to investigate the relationship between predictor variables and BT. The dependent variable was BT or iron deficiency anemia and independent variable by Code 1 was related to BT status.
Therefore, adjusted OR values report the chance of having BT. Logistic regression has some assumptions to run the model. It is necessary to control the amount of multicollinearity between independent variables in multiple regression. Thus, all of the afore-mentioned variables (Table 1) were entered into the model and the amount of multicollinearity was reported (Table 2). Multicollinearity in some of the variables was more than 10 and it is necessary to correct the model.

Table 2: Results of Full Model with Multicollinearity

Index	*Multicollinearity*
RBC	39.769
HGB	558.607
HCT	407.682
MCV	144.536
MCH	283.580
RDW	1.955
HbA2	1.489
Ferritin	1.523
HbA	1.471

To solve the problem of multicollinearity, we tried to select independent variables with the Backward Stepwise: Likelihood Ratio approach. The process of selecting independent variables was completed in the fourth step with HGB, HCT, MCV, MCH, HbA2, and HbA variables. The accuracy of the model classification was 97.1% (Nagelkerke R-square=86.2). Then, the presence of multicollinearity was investigated again, and high multicollinearity between HGB, HCT, MCV, and MCH variables was reported once more.

In the next step, to survey the relationship between variables, a correlation was estimated. For this purpose, we calculated Spearman's correlation. A very high and significant correlation was found between MCV with MCH (r_s=0.75 (<0.001)) and HGB with HCT (r_s=0.89 (<0.001)). Finally, HCT and MCH were omitted from the model, and multicollinearity was maintained in the advised amount.

The final model was fitted with 173 items including HGB, MCV, HbA2, and HbA variables. The goodness of fit for the final model was confirmed by Omnibus Tests of Model Coefficients that was reported to 91.495 (*P*<0.001). Nagelkerke R-square index and Hosmer and Lemeshow Test were equal to 70% and 3.851 (*P*=0.870), respectively. The sensitivity, specificity, and accuracy of the model were estimated as 97%, 72%, and 93%, respectively.

Table 3: Results of final logistic regression fitted to variables

Index	*B*	*S.E.*	*Wald*	*df*	*Sig.*	*Exp(B)*
HGB	0.069	0.205	0.113	1	0.736	1.071
MCV	0.051	0.079	0.412	1	0.521	1.052
HbA2	2.145	0.511	17.656	1	0.000	8.544
HbA	0.019	0.026	0.527	1	0.468	1.019
Constant	-15.896	8.249	3.714	1	0.054	0.000

Table 3 reported that the effect of HbA2 variable was significant (*P*<0.001). The size of the OR index was equal to 8.544 which means that each unit increase in HbA2 increases the chance of developing beta-thalassemia by 8.5 times. Next, the ROC curve is also drawn (Fig. 2). The area under the ROC curve (0.95 CI) was reported as 0.84 (0.94-0.74).

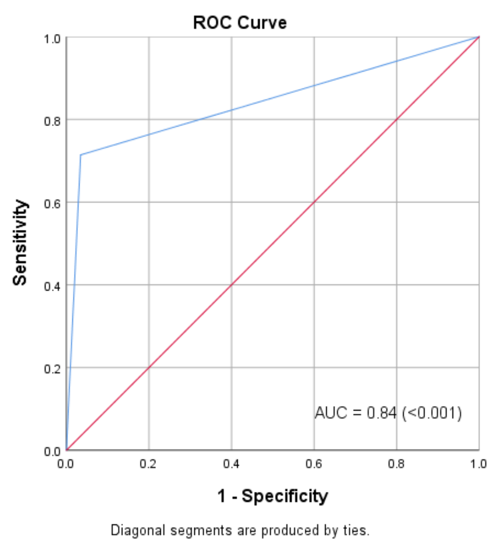

Fig. 2: ROC curve with area under curve (Sig.)

Logistic regression formula to differentiate patient with BTT from IDA is reported in equation 1. This formula will help physicians to guess about nature of anemia (Pr. means Probability).

$$Pr(Patient\ has\ BTT) = \frac{\exp(-15.896 + 0.069*HGB + 0.051*MCV + 2.145*HBA2 + 0.019*HbA)}{1+\exp(-15.896 + 0.069*HGB + 0.051*MCV + 2.145*HBA2 + 0.019*HbA)}$$
[1]

Finally, logistic regression with HbA2 variable was fitted and the accuracy of the model was estimated at approximately 91%. Therefore, if we intend to distinguish BT trait from iron deficiency anemia with only HbA2 variable, we can check it with 91% accuracy.

Discussion

In this study, the data of 292 couples referred to the BT center were used to assess the factors affecting the differentiation of BT from IDA. Various studies have been conducted in this field but high accuracy along with the simplicity of the analysis method is needed. A complex neural network algorithm used to identify important factors in the diagnosis of BTT from IDA (12). The specificity, sensitivity, and accuracy of their study are reported as 92.33%, 93.13%, and 92.5%, respectively, which are similar (97%, 72%, and 93%,) to the results of the present study with a simple regression method. Our study proposed a regression formula that potentially can be used by physicians to differentiate BTT from IDA. In the afore-mentioned study (12), blood indices such as RBC, HGB, MCV and MCH were used to fit the model since RBC and MCV Indices were highly linear with HGB and MCH variables, respectively. Multi collinearity between variables affected the results. In my opinion and based on multicollinearity results, the presence of one of them is sufficient and it is not necessary to use all the afore-mentioned variables.

In another study (14), instead of fitting regression or statistical models, the equations in the published studies were used for differential analysis. The accuracy for all of the twenty equations was calculated with a unique formula, the ROC curve was drawn, and the amount of AUC was estimated. The blood indices used in each of the equations are dependent on that equation. In this study, the most accurate formulas named Bordbar, Kerman I, Kerman II, and Srivastava have an accuracy of 0.86, 0.68, 0.67, and 0.67, respectively, which were lower than the accuracy of the present study (93%). Each equation or formula can be applied to target population since characters of every population is particular.

A study (15) was conducted on 1300 unmarried Indian women (antenatal women) including 66 cases with BT. According to this study, about 1%-17% and an average of 3.3% of the people in this country are BT carriers. The main purpose of the study was to determine the cut-off point of MCV and MCH indices. For this purpose, sensitivity and Yoden index, and ROC curve were used. The cut-off point determined for each index had poor specificity despite high sensitivity. At the cut-off point for MCV and MCH equal to 80fl and 27pg, respectively, the sensitivity and specificity were estimated to be 89.4% and 16.4%, respectively. The AUC for MCV and MCH was reported as 0.65 and 0.635, respectively, and they concluded that although MCV blood index has a better performance than MCH, both indices have very low specificity (15). High-level simultaneous specificity and sensitivity, not just one of them, is a good feature.

Since in the present study, high sensitivity and specificity reported, this study introduces some more advantages compared to the study (15) who determined specific cut-off point for MCV and MCH. They then compared indices of this study with similar, may be incorrect.

A study (16) was conducted on 454 pregnant women with anemia, including 340 women with IDA, 66 women with BT trait, and 48 women with alpha thalassemia. Firstly, the performance of blood differential indices checked using the usual cut-off point, and then a new cut-off point was reported for some of these indices. In addition, two new differential formulas have been developed. This study has reported is a statistically significant difference between the size of RBC, MCV, and MCH index in people with IDA, BT trait and alpha thalassemia, but the afore-mentioned relationship was not significant for HGB and HCT indices. The authors of this study also investigated the potential ability of each blood index to differentiate BT from IDA. The report showed that the sensitivity and specificity of relevant indices such as RBC, HGB, HCT, MCV, MCH, in differentiating BTT from IDA were not acceptable, which means that despite the high accuracy of one of the afore-mentioned accuracy indices, another index was low. For example, although HGB has high sensitivity (89.41), the specificity of this blood index was very low (20.83). Two new formulas (MHA1 and MHA2) were introduced also (16). Each of these formulas was a combination of common blood indices. Therefore, the ability of the previous differential formulas was measured based on the standard values of blood indices as well as with the modified values in the afore-mentioned article. The sensitivity and specificity of the MHA1 formula (MHA2) in differentiating BT from IDA were reported as 83.33 and 65.59 (72.9 and 89.7), respectively. The sensitivity of the first formula and the specificity of the second formula were acceptable. The conclusion was that the sensitivity and specificity of Xiao et al study (16) were lower than our study. Our study indicated that mean of blood indices are statistically different between IDA and BTT groups unless HCT. Finally, in the present study, regression differentiation formula was proposed with four indices. A physician by replacing the amount of every index in the formula can decide about type of patients' anemia.

Differentiating formulas and different methods have been proposed to differentiate BTT from IDA. In the study (13), the differential value of more than 20 formulas has been measured. Other studies have used simple to complex statistical methods for the relevant purpose. In each of the studies, accuracy indices have been used to evaluate the innovative formula. Logistic regression was used to find the differential formula with the highest sensitivity, specificity, and other accuracy measurements (13). To achieve accurate results, logistic regression assumptions, especially multicollinearity, were checked and necessary corrections applied.

Conclusion

It is suggested to use HGB, MCV, HbA and especially HbA2 blood indices to differentiate BT from IDA. The sensitivity, specificity, and accuracy of this formula was estimated to be very good, which was more exact than similar studies. Our study proposed a simple differentiation formula (equation 1). The result of this study can be a warning to the families of married couples to know if their children have the characteristics extracted from this study; at least, until the definitive answer from afore-mentioned center, they can postpone their children's marriage. For future studies, it is suggested to use another statistical method such as Cluster Analysis.

Journalism Ethics considerations

Ethical issues (Including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, redundancy, etc.) have been completely observed by the authors.

Acknowledgements

This study was not funded any way.

Conflict of interest

The authors declare that there is no conflict of interests.

References

1. CDC. What is Thalassemia? National Center on Birth Defects and Developmental Disabilities, Centers for Disease Control and Prevention, New York.
2. Modell B, Darlison M (2008). Global epidemiology of haemoglobin disorders and derived service indicators. *Bull World Health Organ,* 86 (6):480-7.

3. Yolanda S (2021) Thalassemia Prevalence. News-Medical.
4. Khodaei GH, Farbod N, Zarif B, et al. (2013). Frequency of Thalassemia in Iran and Khorasan Razavi. *Int J Pediatr,* 1 (1):45-50.
5. Miri M, Tabrizi Namini M, Hadipour Dehshal M, et al (2013). Thalassemia in Iran in Last Twenty Years: the Carrier Rates and the Births Trend. *Iranian Journal of Blood and Cancer,* 6 (1):11-17.
6. Hashemieh M, Timori Naghadeh H, et al (2015). The Iran Thalassemia Prevention Program: Success or Failure? *Iran J Ped Hematol Oncol,* 5 (3):161-166.
7. Jahangiri M, Rahim F, Saki Malehi A, et al. (2020). Differential Diagnosis of Microcytic Anemia, Thalassemia or Iron Deficiency Anemia: A Diagnostic Test Accuracy Meta-Analysis. *Mod Med Lab J,* 3 (1):16-29.
8. Jameel T, Baig M, Ahmed I, et al. (2017). Differentiation of beta thalassemia trait from iron deficiency anemia by hematological indices. *Pak J Med Sci,* 33 (3):665-669.
9. Jaripour ME, Hayatigolkhatmi K, Iranmanesh V, et al (2018). Prevalence of β-Thalassemia Mutations among Northeastern Iranian Population and their Impacts on Hematological Indices and Application of Prenatal Diagnosis, a Seven-Years Study. *Mediterr J Hematol Infect Dis,* 10 (1):e2018042.
10. Joosten E (2018). Iron deficiency anemia in older adults: A review. *Geriatr Gerontol Int,* 18 (3):373-379.
11. Arora S, Rana D, Kolte S, et al. (2018). Validation of new indices for differentiation between iron deficiency anemia and beta thalessemia trait, a study in pregnant females. *Int J Sci Rep,* 4:26.
12. Kabootarizadeh L, Jamshidnezhad A, Koohmareh Z (2019). Differential Diagnosis of Iron-Deficiency Anemia from Î²-Thalassemia Trait Using an Intelligent Model in Comparison with Discriminant Indexes. *Acta Inform Med,* 27 (2):78-84.
13. Kutner M.H, Nachtsheim C.J, Neter J, W L (2004). ALSM. In. Ed(s): McGraw-Hill Irwin, pp. 409.
14. Ebrahimpour Sadagheyani H, Sharafkhani R, Sakhaei S, et al. (2022). The Evaluation of Results of Twenty Common Equations for Differentiation of Beta Thalassemia Trait from Iron Deficiency Anemia: A Cross-Sectional Study. *Iran J Public Health,* 51(4):929-938.
15. Baliyan M, Kumar M, Nangia A, Parakh N (2019). Can RBC Indices be Used as Screening Test for Beta-Thalassemia in Indian Antenatal Women? *J Obstet Gynaecol India,* 69 (6):495-500.
16. Xiao H, Wang Y, Ye Y, et al (2021). Differential diagnosis of thalassemia and iron deficiency anemia in pregnant women using new formulas from multidimensional analysis of red blood cells. *Ann Transl Med,* 9 (2):141.