



**Original Article**

## Detection of $\beta$ -Thalassemia Carriers by Red Cell Parameters Obtained from Automatic Counters using Mathematical Formulas

Idit Lachover Roth<sup>1</sup>, Boaz Lachover\*, Guy Koren\*\*, Carina Levin<sup>2,3</sup>, Luci Zalman<sup>4</sup> and Ariel Koren<sup>2,3</sup>.

<sup>1</sup> Allergy and Clinical Immunology Unit, Meir Medical Center, Kfar Saba, Israel.

<sup>2</sup> Pediatric Hematology Unit, Emek Medical Center, Afula, Israel.

<sup>3</sup> The Ruth and Baruch Faculty of Medicine, Technion – Israel Institute of Technology, Haifa, Israel.

<sup>4</sup> Hematology Laboratory, Emek Medical Center, Afula, Israel.

\* Passed away before manuscript submission.

\*\* Personal collaboration.

**Competing interests:** The authors have declared that no competing interests exist.

**Abstract. Background:**  $\beta$ -thalassemia major is a severe disease with high morbidity. The world prevalence of carriers is around 1.5–7%. The present study aimed to find a reliable formula for detecting  $\beta$ -thalassemia carriers using an extensive database of more than 22,000 samples obtained from a homogeneous population of childbearing age women with 3161 (13.6%) of  $\beta$ -thalassemia carriers and to check previously published formulas.

**Methods:** We applied a mathematical method based on the support vector machine (SVM) algorithm in the search for a reliable formula that can differentiate between thalassemia carriers and non-carriers, including normal counts or counts suspected to belong to iron-deficient women.

**Results:** Shine's formula and our SVM formula showed >98% sensitivity and >99.77% negative predictive value (NPV). All other published formulas gave inferior results.

**Conclusions:** We found a reliable formula that can be incorporated into any automatic blood counter to alert health providers to the possibility of a woman being a  $\beta$ -thalassemia carrier. A further simple hemoglobin characterization by HPLC analysis should be performed to confirm the diagnosis, and subsequent family studies should be carried out. Our SVM formula is currently limited to women of fertility age until further analysis in other groups can be performed.

**Keywords:** Thalassemia; Prevention; Screening; Mathematical formulas.

**Citation:** Roth I.L., Lachover B., Koren G., Levin C., Zalman L., Koren A. Detection of  $\beta$ -thalassemia carriers by red cell parameters obtained from automatic counters using mathematical formulas. *Mediterr J Hematol Infect Dis* 2018, 10(1): e2018008, DOI: <http://dx.doi.org/10.4084/MJHID.2018.008>

**Published:** January 1, 2018

**Received:** September 10, 2017

**Accepted:** December 8, 2017

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by-nc/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Correspondence to: Ariel Koren MD. Pediatric Hematology Unit, Emek Medical Center, Afula, Israel. Tel: +972-4-6495615, Fax: +972-4-6495237. E-mail: [koren\\_a@clalit.org.il](mailto:koren_a@clalit.org.il)

**Introduction.**  $\beta$ -thalassemia is considered the world's most widespread genetic disease. Between 1.5 and 7% of the world population carries one of the genes that cause hemoglobinopathies, and about 60,000 a year are diagnosed as  $\beta$ -thalassemia patients.<sup>1,2</sup> In Israel, the incidence of

$\beta$ -thalassemia carriers is around 20% among Kurdish Jews and between 5 and 10% among the Arab population.<sup>3</sup> Thalassemia patients require regular blood transfusions and suffer from severe iron overload, which is the main cause of morbidity and mortality. Despite an actual

treatment, which needs significant medical and financial resources,<sup>4</sup> quality of life and life expectancy are still lower than in the general population. The disease's severity, high cost and high prevalence in developing countries with low financial capability justify the implementation of prevention programs in those countries.

Different strategies are implemented to cope with social and religious beliefs. In Israel, a screening program was initiated in 1987.<sup>5</sup> The blood count in  $\beta$ -thalassemia carriers shows low mean corpuscular volume (MCV) and low mean corpuscular hemoglobin (MCH). These parameters, which are easily measured by automated blood cell counters, can indicate suspicion of a carrier state. The MCV and MCH are in the same range in  $\beta$ -thalassemia carriers and patients with iron deficiency anemia (IDA), but the red blood cell (RBC) count and red cell distribution width (RDW) can differentiate between the two.

The gold standard for the diagnosis of  $\beta$ -thalassemia carriers is electrophoresis or HPLC analysis of hemoglobin (Hgb). Automated blood count results that suggest  $\beta$ -thalassemia carrier status can significantly improve the recognition of carriers and consequently of couples at risk. Those couples can be referred for further genetic counseling. A reliable and inexpensive method for mass screening of the population is needed to enable the selection of samples for further HPLC analysis to confirm the diagnosis.

Several investigators have tried to use blood count parameters for the detection of suspected  $\beta$ -thalassemia carriers compared to patients with IDA or normal individuals by determining cutoff points.<sup>6</sup> The first attempts to use mathematical formulas were made in the early 1970s.<sup>7-9</sup> Those studies were performed on small samples. The only published large-scale research was performed by Shine and Lal in 1977.<sup>10</sup> They developed a new formula and checked its reliability on 25,000 blood samples which included a small sample of only 138  $\beta$ -thalassemia carriers (0.55%).<sup>10</sup> In the last few decades, more attempts have been made using conventional mathematical methods to find a reliable formula, but with no success.<sup>11-35</sup> None of the published formulas are currently in use in daily practice because they were not proven to be reliable enough for routine screening and detection of thalassemia carriers. A summary of the most common published formulas is presented in **Table 1**.

A sample used to validate a mathematical formula for inclusion in routine use needs to be sufficiently large to be considered reliable. Very high specificity and minimal false negative results are required to confirm the formula's reliability. The aim of the present study was to find a reliable formula using a large database of more than 22,000 samples obtained from a relatively homogeneous population of childbearing age woman with a relatively large percentage of  $\beta$ -thalassemia carriers, 3161 samples, (13.6%). The secondary aim was to use our extensive database to check and potentially validate the previously published formulas. We found an innovative formula that can be included as part of the routine analysis in automated blood counters and will give an automatic warning when a  $\beta$ -thalassemia carrier is suspected.

## Methods.

*Study population.* As part of the screening program for  $\beta$ -thalassemia, routinely implemented by the Israel Ministry of Health since 1987, blood samples are collected from all childbearing age women in the northeast of Israel.<sup>5</sup> The ethnic origin of the population covered by this program is equally distributed between Jews and Arabs.

*Laboratory tests.* Blood samples were analyzed by blood count and by Hgb electrophoresis (until 1999) and HPCL (afterward). HPLC analysis included determination of Hgb A, A<sub>2</sub>, and F as well as other pathological hemoglobin.

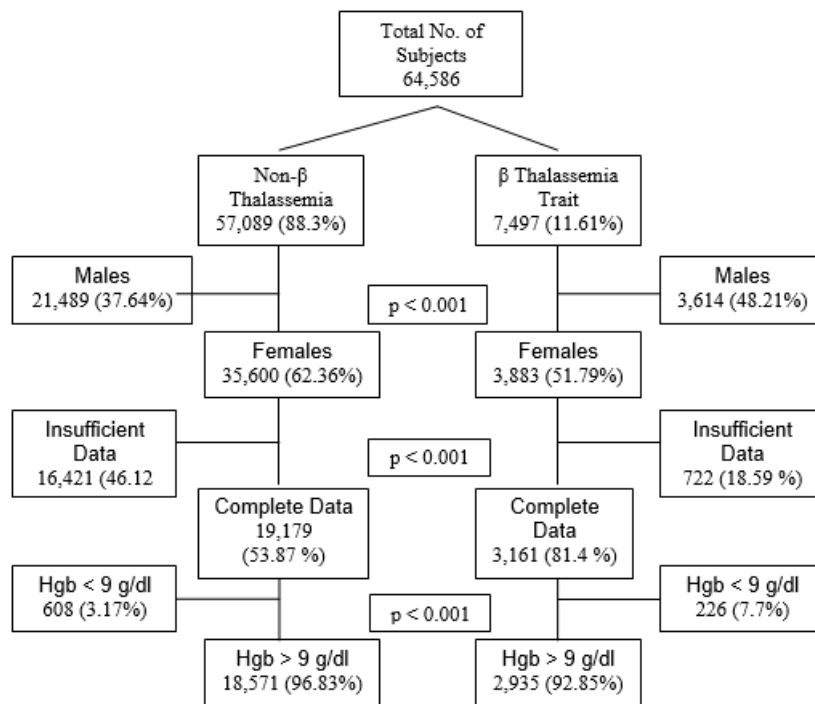
All the samples were analyzed at the same laboratory at the Emek Medical Centre. From the beginning of the survey and until the early nineties the analyses were performed using the Technicon H1 or H2 machine, and from this time until now the Siemens Advia 2120i Hematology System was used. Throughout all the years the hematology laboratory has always had a regular external quality control and qualified by the ISO certificate. We only included blood samples from childbearing age women to establish a homogeneous population. Just  $\beta$ -thalassemia carriers confirmed by Hgb electrophoresis or HPLC and hemograms with normal Hemoglobin electrophoresis or HPLC were included in the analysis. Blood counts with other abnormal hemoglobin results were excluded. The whole database included more than 80,000 samples.

Our data included all samples collected as part of the  $\beta$ -thalassemia screening program from 1987

**Table 1.** Differentiation between  $\beta$ -thalassemia Trait and normal individuals or iron-deficient subjects, summary of the formulas published in the literature.

No. Ref	Study (year)	Formula	$\beta$ TM cutoff	Study size		Authors' Comments/Conclusions	Other references that checked this formula
				Total	$\beta$ TM		
(7)	Srivastava (1973)	$\frac{MCH}{RBC}$	<3.8	N.A.	N.A.	9 out of 10 $\beta$ TM are below the cutoff	(15,17-19,21-23, 25, 32, 33, 35, 41, 42)
(8) (11)	1) England & Fraser (1973) 2) England & Fraser (1979)	MCV-RBC-(5-Hb)-K*	<0	1) 62 2) 462	1) 38 2) 242	1) Not applicable in pregnancy. 6.6% overlap 2) Not suitable for screening for $\beta$ TM	(15,17-20,22-25, 29, 33, 35, 41, 42)
(9)	Mentzer (1973)	$\frac{MCV}{RBC}$	<13	103	50 ( $\alpha$ or $\beta$ )	Hgb < 90 g/l excluded 84% classified correctly 83% classified correctly for England & Fraser index	(8,15,17-19,21-25, 32, 33, 35, 41, 42)
(10)	Shine & Lal (1977)	$\frac{MCV^2 \times MCH}{100}$	<1530	25,302	138	0.7% false negative, 4.4% false positive. 8 to 45% false negative for previous formulas	(8, 9, 15, 17, 19, 21-23, 25, 32, 33, 35, 41, 42)
(12)	Ricerca et al. (1987)	$\frac{RDW}{RBC}$	<3.3	398 (100 normal)	134	98% correct detection of $\beta$ TM	(17, 22, 23, 25, 33, 35, 41)
(13)	Green & King (1989)	$\frac{MCV^2 \times RDW}{(Hb \times 100)}$	<65	135	33	Subjects with normal blood count and childbearing age women excluded	(16, 17, 20, 22-25, 29, 33, 35, 41)
(14)	D'Onofrio et al. (1992)	$\frac{MCV}{MCH}$	<0.9	291	172	Sensitivity 94%, specificity 92.3%, predictive value 94% for IDA	(22, 33, 34)
(43)	Romero Artaza et al. (1999)	$\frac{RDW \times MCV}{RBC}$	<220	90	20	RDW is more sensitive than MCV for establishing diagnosis	(17, 22-24, 33, 41, 44)
(17)	Sirdah et al. (2008)	MCV-RBC-(3XHb)	<27	2196	1272	This formula, Green & King formula and RDW were most reliable for differentiating $\beta$ TM from IDA	(19-22, 29, 32-34, 44)
(18)	Ehsani et al. (2009)	MCV-(10 x RBC)	<15	284	154	The new index correctly identified 92.96% of the patients, inferior only to Mentzer index	(17, 19, 21, 22, 32, 33, 41)
(19)	Sirachainan et al. (2014)	1.5 x Hb - 0.05 x MCV	<14	345	111	Including Hgb E, $\alpha$ or $\beta$ Sensitivity 84.6% Specificity 87.5%	
(32)	Bordbar et al. (2015)	$ 80 - MCV  \times  27 - MCH $	>44.7 6	504	151	Best formulas Shine & Lal and new formula: 87.6%, 84.7% sensitivity and 84.7%, 87.9% specificity, respectively	
(35)	Rahim & Keikhaei (2009)	$\frac{Hb \times RDW \times 100}{RBC^2 \times MCHC}$	<21	323	153	Shine & Lal accurate in children and RBC and RDW in older people	(32, 33)
(33)	Hisham index (2015)	$\frac{MCH \times RDW}{RBC}$	<67	600 (with microcytic anemia)	372	Sensitivity for published indexes 25%-97% $\beta$ TM specificity: Hisham - 92%, Hameed - 91%	
(33)	Hameed index (2015)	$\frac{MCH \times Hct \times RDW}{(RBC \times Hb)^2}$	<220				
(26)	Amendolia et al. - SVM (2003)	Single Vector Analysis (RBC, Hgb, Hct, MCV chosen empirically)		304	81 $\alpha$ 27 $\beta$	83% sensitivity and 95% specificity. ID excluded 15% false negative	

$\beta$ TM:  $\beta$ -thalassemia Trait, IDA: iron deficiency anemia, ID: iron deficiency, RBC: red blood count, Hgb: hemoglobin, Hct: hematocrit, MCV: mean corpuscular volume, MCH: mean corpuscular hemoglobin, MCHC: mean corpuscular hemoglobin concentration; RDW: red cell distribution width, N.A.: not applicable.



**Figure 1.** Total number of subjects initially included in the study and final number of subjects analyzed. \*Samples with all red cell index values (Hgb, Hct, RBC, MCV, MCH, MCHC and RDW).

to 2013. As shown in **Figure 1**, 64,586 samples were initially included in the study. 57,089 blood counts were from healthy individuals in whom the  $\beta$ -thalassemia trait was ruled out by hemoglobin electrophoresis or HPLC analysis. The second group of 7497 samples belongs to proven  $\beta$ -thalassemia trait. After selection of only women and samples for which all red cell indices were included, there were 22,340 samples in our study, including 19,179 (85.85%) samples from healthy women or women with IDA (non- $\beta$ -thalassemia cohort) and 3,161 (13.6%) samples from female  $\beta$ -thalassemia carriers.

Insufficient data for SVM analysis were found in 17,143 samples.

From the 22,340 samples, 21,506 had Hgb level above 90 g/l, and of these, 18,571 (86.35%) belonged to the non- $\beta$ -thalassemia group and 2935 (13.64%) to the  $\beta$ -thalassemia carriers. The proportion of  $\beta$ -thalassemia carriers was higher in the group of women with Hgb level below 90 g/l ( $p < 0.001$ ).

Since the combination of being a  $\beta$ -thalassemia carrier and having IDA, a frequent state in childbearing age women, can cause more severe anemia than each condition alone, we analyzed the data from blood counts with Hgb level above or

below 90 g/l separately. Plasma Iron or ferritin levels were not routinely analyzed as part of the screening program.

*Mathematical formulas.* We applied the blood count parameters of our collected data to the formulas published in the literature. After this analysis, we used another mathematical method, based on the support vector machine (SVM) algorithm, to the same data with the aim of finding a new and reliable formula that can differentiate between  $\beta$ -thalassemia carriers and non-carriers, including normal counts and counts suspected of stemming from women with IDA. Briefly, this SVM algorithm intends to find the linear differentiation between two categories; in the first step the algorithm uses measurable values for the two categories and calculates the maximal distance that can differentiate between the two categories with a minimal overlapping of the two series. If a linear differentiation cannot be found, then the algorithm intends to find a nonlinear differentiation. A more detailed explanation is beyond the purpose of this paper and can be found in detail in the published literature (26, 36). The SVM algorithm initially used all of the relevant data from the RBC counts, including Hgb, Hc,

RBC, MCV, MCH, MCHC, and RDW. After running the whole database, the algorithm chooses

the most relevant values and discards the others. The cutoff of the SVM formula is "0". Any

negative number is considered as  $\beta$ -thalassemia carrier, and any positive number is considered as a "noncarrier." We used the SVM method to the samples where all the RBC indexes were available (Figure 1), and applied the same parameters calculated for the previously published formulas. For those formulas the originally published cutoff was used too.

**Statistical methods.** Student *t*-test and chi-square test ( $\chi^2$ ) were used to calculate the differences between the  $\beta$ -thalassemia carriers and non-carriers.  $p < 0.05$  was considered significant.

This study was approved by the local ethics committee.

**Results.** All of the RBC parameters differed significantly between the  $\beta$ -thalassemia carriers and non-carriers. However, for the data obtained from both groups with Hgb  $< 90$  g/l, the differences between the RDW values were not significant, probably due to the presence of

combined  $\beta$ -thalassemia carriers and IDA in both groups (Table 2).

The formulas already published in the literature (Table 1) were applied to our whole database, and the results are given in Table 3 for Hgb  $> 90$  g/l and in Table 4 for Hgb  $< 90$  g/l. The SVM equation that finally used the MCV and MCH values corrected with different constants is shown.

**Discussion.** For the physician, RBC indices should be sufficient to raise suspicion of a  $\beta$ -thalassemia carrier and therefore to perform a further evaluation. Despite this logical rationale, most  $\beta$ -thalassemia carriers are only diagnosed in population-screening programs or when a new family case is discovered. This reasoning and the burden of  $\beta$  thalassaemic patients for health services have pushed many countries to develop effective screening programs for  $\beta$ -thalassemia prevention.<sup>37</sup>

Since the early 1970s, many investigators have tried to find a reliable formula for  $\beta$ -thalassemia

**Table 2.** Red cell values for non- $\beta$ -thalassemia and  $\beta$ -thalassemia Trait women.

Group		Hgb $> 90$ g/l (N = 21506)		Hgb $< 90$ g/l (N = 834)	
		Non $\beta$ TM (N = 18571)	$\beta$ TM (N = 2935)	Non $\beta$ TM (N = 608)	$\beta$ TM (N = 226)
RBC ( $\times 10^9$ /dl)	Average $\pm$ STD	4.2 $\pm$ 0.43	5.42 $\pm$ 0.55	3.93 $\pm$ 0.62	4.55 $\pm$ 0.58
	Range	2.17 – 7.67	3.21 – 7.81	1.4 – 6.23	3.06 – 6.23
	p value (*)	$< 0.001$		$< 0.001$	
Hgb (gr/dl)	Average $\pm$ STD	117.5 $\pm$ 10.6	106.5 $\pm$ 9.5	79.4 $\pm$ 9.3	83.9 $\pm$ 5.9
	Range	90 – 192	90 – 154	33 – 89	60 – 90
	p value (*)	$< 0.001$		$< 0.001$	
Hct (%)	Average $\pm$ STD	35.89 $\pm$ 3.21	34.04 $\pm$ 3.29	27.11 $\pm$ 3.2	27.52 $\pm$ 2.3
	Range	10.2 – 57.9	11.3 – 48	12.3 – 35.9	19.6 – 33.4
	p value (*)	$< 0.001$		0.077	
MCV (fl)	Average $\pm$ STD	85.91 $\pm$ 6.8	63.13 $\pm$ 8.75	70.03 $\pm$ 9.28	60.9 $\pm$ 6.63
	Range	54.4 – 125.9	48 – 91.5	48.8 – 117	45 – 83
	p value (*)	$< 0.001$		$< 0.001$	
MCH (pg)	Average $\pm$ STD	28.16 $\pm$ 2.7	19.71 $\pm$ 1.94	20.6 $\pm$ 3.68	18.7 $\pm$ 2.39
	Range	16.2 – 40.7	14 – 31	13.5 – 39.4	12.8 – 28.3
	p value (*)	$< 0.001$		$< 0.001$	
MCHC (gr/dl)	Average $\pm$ STD	32.75 $\pm$ 1.76	31.22 $\pm$ 1.59	29.37 $\pm$ 2.16	30.5 $\pm$ 2.03
	Range	12.6 – 41.2	17.5 – 39	18.1 – 39	24 – 39
	p value (*)	$< 0.001$		$< 0.001$	
RDW (%)	Average $\pm$ STD	14.95 $\pm$ 1.98	16.31 $\pm$ 1.56	18.76 $\pm$ 2.84	18.2 $\pm$ 2.2
	Range	10.1 – 34	12 – 22.8	12.5 – 32.2	14.6 – 29.5
	p value (*)	$< 0.001$		0.0079	

$\beta$ TM:  $\beta$ -thalassemia Trait, RBC: red blood count, Hgb: hemoglobin, Hct: hematocrit, MCV: mean corpuscular volume, MCH: mean corpuscular hemoglobin, MCHC: mean corpuscular hemoglobin concentration, RDW: red cell distribution width. STD: standard deviation. \*P-value between non  $\beta$ TM and  $\beta$ TM within the same Hgb range group ( $< 90$  g/l or  $> 90$  g/l).

**Table 3.** Calculation of the results of the present data for the different published formulas. Subjects with Hgb > 90 g/l.

Formula (Year)	False Positive (%) [N <sup>o</sup> ]	False Negative (%) [N <sup>o</sup> ] **	Sensitivity (%) *	Specificity (%) *	PPV (%) *	NPV (%) *	Youden Index
Srivastava (1973)	0.66 [122]	37 [1091]	62.84	<b>99.34</b>	<b>93.79</b>	<b>94.41</b>	62.18
England & Fraser (1973)	2.47 [458]	26.3 [773]	73.67	<b>97.53</b>	82.52	<b>95.9</b>	71.2
Mentzer (1973)	1.02 [189]	24.4 [718]	75.54	<b>98.98</b>	92.14	<b>96.24</b>	74.52
Shine & Lal (1977)	11.77 [2187]	<b>1.26 [37]</b>	<b>98.74</b>	88.22	56.99	<b>99.77</b>	86.96
Ricerca (1987)	31.67 [5882]	25.4 [746]	74.59	68.33	27.13	<b>94.44</b>	42.92
Green & King (1989)	19.48 [3618]	<b>7.9 [233]</b>	<b>92.06</b>	80.52	42.76	<b>98.46</b>	72.58
D'Onofrio (1992)	100 [18572]	<b>0 [0]</b>	N.A.	N.A.	N.A.	N.A.	N.A.
Romero Artaza (1999)	2.6 [482]	20 [593]	79.8	<b>97.4</b>	82.93	<b>96.82</b>	77.2
Sirdah (2008)	0.45 [83]	38 [1134]	61.38	<b>99.55</b>	95.59	<b>94.22</b>	60.93
Ehsani (2009)	0.92 [170]	25 [732]	75.07	<b>99.08</b>	92.83	<b>96.17</b>	74.15
Sirachainan (2014)	67.5 [12534]	19 [558]	80.99	32.51	15.94	<b>91.54</b>	13.5
Bordbar (2015)	15.37 [2854]	<b>9.4 [278]</b>	<b>90.53</b>	84.63	48.22	<b>98.26</b>	75.16
Rachim (2009)	1.29 [240]	28.1 [825]	71.9	<b>98.71</b>	89.79	<b>95.69</b>	70.61
Hisham index (2015)	1.6 [298]	25 [734]	75	<b>98.4</b>	88.08	96.13	73.4
Hameed index (2015)	100 [18572]	<b>0 [0]</b>	N.A.	N.A.	N.A.	N.A.	N.A.
SVM (Present Study)	11.82 [2196]	<b>1.1 [33]</b>	<b>98.88</b>	88.18	56	<b>99.7</b>	86.99

\*Values above 90% are presented in Bold. \*\*Values below 10% are presented in Bold. PPV: positive predictive value, NPV: negative predictive value. Hgb: hemoglobin. SVM: single vector machine. N.A.: not applicable.

**Table 4.** Calculation of the results of the present data for the different published formulas. Subjects with Hgb<90g/l.

Formula (Year)	False Positive (%) [N <sup>o</sup> ]	False Negative (%) [N <sup>o</sup> ] **	Sensitivity (%) *	Specificity (%) *	PPV (%) *	NPV (%) *	Youden Index
Srivastava (1973)	10.86 [66]	68 [155]	31.42	89.14	51.82	77.76	20.56
England & Fraser (1973)	0.82 [5]	76 [172]	23.89	99.18	91.52	77.8	23.07
Mentzer (1973)	7.24 [44]	60 [136]	39.82	<b>92.76</b>	67.16	80.57	32.58
Shine & Lal (1977)	77.96 [474]	<b>0.44 [1]</b>	<b>99.56</b>	22.04	32.18	<b>99.25</b>	21.6
Ricerca (1987)	0 [0]	89 [203]	10.18	<b>100</b>	<b>100</b>	74.96	10.18
Green & King (1989)	4.44 [27]	47 [107]	52.65	<b>95.56</b>	81.5	84.44	48.21
D'Onofrio (1992)	100 [608]	<b>0 [0]</b>	N.A.	N.A.	N.A.	N.A.	N.A.
Romero Artaza (1999)	0.66 [4]	68 [154]	31.86	<b>99.34</b>	94.73	79.68	31.2
Sirdah (2008)	2.14 [13]	70 [159]	29.65	<b>97.86</b>	83.75	78.91	27.51
Ehsani (2009)	9.38 [57]	56 [128]	43.36	<b>90.63</b>	63.22	81.14	33.99
Sirachainan (2014)	100 [608]	<b>0 [0]</b>	N.A.	N.A.	N.A.	N.A.	N.A.
Bordbar (2015)	69.24 [421]	<b>4.9 [11]</b>	<b>95.13</b>	30.76	33.8	<b>94.44</b>	25.89
13 Rachim (2009)	2.14 [13]	72 [163]	27.88	<b>97.86</b>	82.89	78.49	25.74
Hisham index (2015)	6.41 [39]	66.8 [151]	33.19	<b>93.59</b>	65.78	79.02	26.78
Hameed index (2015)	100 [608]	<b>0 [0]</b>	N.A.	N.A.	N.A.	N.A.	N.A.
SVM (Present Study)	85.2 [518]	<b>0.44 [1]</b>	<b>99.56</b>	14.8	30.28	<b>98.9</b>	14.36

\*Values above 90% are presented in bold. \*\* Values below 10% are presented in bold. PPV: positive predictive value, NPV: negative predictive value, Hgb: Hemoglobin, SVM: single vector machine, N.A.: not applicable.

detection based on RBC indices, but none of those formulas had been ever integrated as part of the

routine blood counts in automated machines. All of the studies, except one, consisted of analyses of

small numbers of patients (maximum of 2,196 samples). The only study that analyzed a large database was published in 1977 by Shine and Lal,<sup>10</sup> with 25,302 samples, but only 138 (0.55%)  $\beta$ -thalassemia carriers.

Kiss et al.<sup>27</sup> combined low MCV and ethnic background to create an algorithm that predicts the probability of detecting  $\beta$ -thalassemia carriers. Their database included 789 patients with MCV < 80 fl and only 31 patients diagnosed as  $\beta$ -thalassemia carriers. They concluded that by using MCV and ethnic background it was sufficient to detect  $\beta$ -thalassemia carriers, but that study was performed in a population where  $\beta$ -thalassemia is prevalent in a specific ethnic community and might not be applicable when the incidence of  $\beta$ -thalassemia carriers is dispersed throughout the whole population.<sup>27,28</sup> Amendolia et al.<sup>26</sup> were the first to use separation algorithms, including SVM, to differentiate  $\alpha$  and  $\beta$ -thalassemia carriers from healthy subjects, but they excluded IDA patients. They found 83% sensitivity with the SVM algorithm. Since the SVM analysis used all the possible relevant parameters, in our study, we exclude from the analysis all the samples where some of the RBC indexes are missing. The calculation for all the other formulas was based on the same cohort that the SVM formula was calculated. Urrechaga et al.<sup>30</sup> used multiple discriminant analysis to differentiate IDA from  $\beta$  and  $\alpha$  thalassemia carriers; they did not intend to separate thalassemia carriers from healthy subjects. Sargolzaie and Miri-Moghaddam<sup>29</sup> recently described the use of binary logistic regression analysis to find the best equation for a group of 100  $\beta$ -thalassemia carriers and 77 IDA patients, in this study healthy subjects were not included. The most recent attempts to find a reliable formula or algorithm were limited to discriminating subjects with microcytic anemia, excluding people with low Hgb and including small control groups.<sup>31,35</sup> In 2015, Bordbar et al.<sup>32</sup> applied five previously published formulas and a new one on the blood count data from 504 patients, 151 of them  $\beta$ -thalassemia carriers. The highest sensitivity and specificity found for detection of  $\beta$ -thalassemia carriers was 87%. A meta-analysis examined the most frequently used indices and concluded that MCV and MCH are the most important ones, but the sensitivity and specificity of the published formulas were not high enough to make a definitive diagnosis of thalassemia carriers.<sup>34</sup>

In a reliable formula, a negative predictive value (NPV) higher than 99% is enough to recognize a formula reliable for daily use. A program that intends to become safe for mass population screening should miss as few false negative samples as possible.

In our study, we applied all of the published formulas to our vast database that included more than 22,000 samples with 3161 (13.6%)  $\beta$ -thalassemia carriers. Only one of these formulas was reliable enough to differentiate between  $\beta$ -thalassemia carriers and healthy or IDA patients: that published by Shine and Lal.<sup>10</sup> This formula showed only 37 false negative results among a total number of 2936 samples, representing a sensitivity of 98.74% and an NPV of 99.77%. Our new SVM formula gave similar results: 34 false negative results, a sensitivity value of 98.84% and NPV of 99.79%. All 34 false negative samples detected with the SVM algorithm were also found to be false negative with Shine and Lal's formula. In all of those false negative results, the MCV was  $\geq 77.6$  fl. A high MCV is known to be characteristic of specific mutations with milder disease.<sup>38</sup> The specificity of those formulas was also high, 88.22% for Shine and Lal's formula and 88.15% for our SVM formula.

The accuracy of the HPLC analysis for detection of  $\beta$ -thalassemia carriers might be limited if the Hgb level is below 90 g/l. For patients with Hgb < 90 g/l, it is recommended to correct the iron deficiency before performing the HPLC analysis. We analyzed the results for a group with Hgb level below 90 g/l that included 834 samples. Among those samples we found a significant reduction of the false negative results, 1/226 in Shine and Lal and in the SVM formula, but we had more cases false positive, which mean that those formulas became more sensitive but less specific. When we applied the Mentzer index to the group with low Hgb, the results found a sensitivity of only 75.54%, a high specificity of 98.98% but an NPV of 96.24, which is lower than the Shine and Lal and SVM formulas.<sup>9</sup> Since the screening program does not include analysis of iron status, we cannot prove which proportion of samples with Hgb level below 90 g/l belong to  $\beta$ -thalassemia carriers or IDA or both.

Although the prevention programs implemented in some countries<sup>5,39</sup> have succeeded to lower the prevalence of giving birth to affected children, they have several limitations, financial and operative. Efficient utilization of the RBC indices'

automatic alerts can increase the number of detected  $\beta$ -thalassemia carriers without the need for special attention from the medical staff. The HPLC test costs more than a routine blood count, and if we can routinely include an automatic alert signal in any routine blood count analyzed by the automated machines, the referring physician or nurse can get an indication of whether they need to perform further studies and genetic counseling.

As far as we know, our study is the first to exam all of the formulas present in the current literature. Our group found a new formula on a numerous database of more than 22,000 samples, including a large group of  $\beta$ -thalassemia carriers, 3161 samples (13.6%) compared to Shine study that included a small percentage of carriers.

Our strategy for  $\beta$ -thalassemia screening is based on choosing childbearing age women. The selection of blood samples from only these women does not reduce the reliability of our formulas or the ability to use them on a routine basis. This choice made our study database much more homogeneous and helped us to obtain more reliable results.

The goal of a reliable screening test is to get as close as possible to zero false negative results with a minimal percentage of false positive results. We did not succeed in getting zero false negative results, but only Shine and Lal's formula<sup>10</sup> and our SVM formula had similar very low false negative results, in a range that is acceptable for population screenings. All of the other published formulas had too many false negative which made them insufficiently reliable as a screening tool. Only Shine and Lal's and our SVM formulas succeeded to detect  $\beta$ -thalassemia carriers with normal, or near normal blood count indices. We succeeded in identifying 1074  $\beta$ -thalassemia carriers with Hgb  $\geq$  110 g/l, 67 with MCV  $\geq$  75 fl and even 44 women who had Hgb  $\geq$  110 g/l and MCV  $\geq$  75 fl.

The finding of 11% false positive results meant that all those suspected of being carriers would require further evaluation by HPLC test. We think that this is a very "small" price to pay considering the number of carriers that we can discover in a routine blood sample by increasing the NPV to

over 99% and lowering the undetected false negative cases (**Table 3**).

Integrating a formula into a routine blood count to detect samples that raise suspicion of being  $\beta$ -thalassemia carriers means that we can add an alert in the test result that says: "suspected  $\beta$ -thalassemia carrier, please refer to HPLC test." In this way, the detection of  $\beta$ -thalassemia carriers will no longer depend only on the alertness of the health-care staff.

The accuracy of the HPLC analysis for blood samples with low Hgb is lower than the results when Hgb is above 90 g/l. In  $\beta$ -thalassemia carriers with IDA, Hgb A<sub>2</sub> levels can be lower than in non-IDA  $\beta$ -thalassemia carriers and even in the normal range, and in patients with severe IDA, it is recommended that the iron deficiency is treated first, and then the Hgb electrophoresis or HPLC test performed.<sup>40</sup> To the best of our knowledge, this is the first study to analyze samples with Hgb below 90 g/l. In this group, the sensitivity was still very high (99.56%) with less than 0.5% of false negative results.

$$\frac{1.449994 * (MCV - 82.80762)}{10.28125} + \frac{0.662983 * (MCH - 27.02007)}{3.904478} + 0.981605$$

Figure 2. Single vector machine algorithm calculated formula.

**Conclusions.** We found a reliable formula that can be incorporated into any automatic blood counter and advice health providers when women are suspected of being  $\beta$ -thalassemia carriers. Our SVM formula is currently limited to women of fertility age until further analysis of other groups can be performed. A similar study should be conducted in a large number of  $\alpha$  thalassemia carriers to prove the reliability of the formula in this group. Of course, the diagnosis of  $\alpha$  thalassemia carriers needs to be confirmed by molecular analysis.

**Acknowledgments.** This paper is dedicated to the memory of Boaz Lachover, Idit Lachover Roth's husband, who was killed in a cycling "hit and run" accident before this study was completed.

## References:

- Modell B, Darlison M. Global epidemiology of haemoglobin disorders and derived service indicators. *Bull World Health Organ.* 2008;86(6):480-7. <https://doi.org/10.2471/BLT.06.036673> PMID:18568278 PMCID:PMC2647473
- De Sanctis V, Kattamis C, Canatan D, Soliman AT, Elsedfy H, Karimi M, et al. beta-Thalassemia Distribution in the Old World: an Ancient Disease Seen from a Historical Standpoint. *Mediterr J Hematol Infect Dis.* 2017;9(1):e2017018. <https://doi.org/10.4084/mjhid.2017.018> PMID:28293406 PMCID:PMC5333734
- Filon D, Oron V, Krichevski S, Shaag A, Shaag Y, Warren TC, et al. Diversity of beta-globin mutations in Israeli ethnic groups reflects recent historic events. *Am J Hum Genet.* 1994;54(5):836-43. PMID:8178823 PMCID:PMC1918256



4. Koren A, Profeta L, Zalman L, Palmor H, Levin C, Zamir RB, et al. Prevention of beta Thalassemia in Northern Israel - a Cost-Benefit Analysis. *Mediterr J Hematol Infect Dis*. 2014;6(1):e2014012. <https://doi.org/10.4084/mjhid.2014.012> PMID:24678389 PMCID:PMC3965716
5. Koren A, Zalman L, Palmor H, Ekstein E, Schneour Y, Schneour A, et al. [The prevention programs for beta thalassemia in the Jezreel and Eiron valleys: results of fifteen years experience]. *Harefuah*. 2002;141(11):938-43, 1210. PMID:12476624
6. Cazzavillan M, Barbui T, Franchi F, Chisesi T, Dini E. Comparison of Price-Jones curves obtained with an electronic corpuscle counter in normal subjects and in patients with thalassaemia and iron deficiency. *Haematologia (Budap)*. 1973;7(3):333-7.
7. Srivastava PC. Differentiation of thalassaemia minor from iron deficiency. *Lancet*. 1973;2(7821):154-5. PMID:4124080
8. England JM, Fraser PM. Differentiation of iron deficiency from thalassaemia trait by routine blood-count. *Lancet*. 1973;1(7801):449-52. [https://doi.org/10.1016/S0140-6736\(73\)91878-3](https://doi.org/10.1016/S0140-6736(73)91878-3)
9. Mentzer WC, Jr. Differentiation of iron deficiency from thalassaemia trait. *Lancet*. 1973;1(7808):882. [https://doi.org/10.1016/S0140-6736\(73\)91446-3](https://doi.org/10.1016/S0140-6736(73)91446-3)
10. Shine I, Lal S. A strategy to detect beta-thalassaemia minor. *Lancet*. 1977;1(8013):692-4. [https://doi.org/10.1016/S0140-6736\(77\)92128-6](https://doi.org/10.1016/S0140-6736(77)92128-6)
11. England JM, Fraser P. Discrimination between iron-deficiency and heterozygous-thalassaemia syndromes in differential diagnosis of microcytosis. *Lancet*. 1979;1(8108):145-8. [https://doi.org/10.1016/S0140-6736\(79\)90532-4](https://doi.org/10.1016/S0140-6736(79)90532-4)
12. Ricerca BM, Storti S, d'Onofrio G, Mancini S, Vittori M, Campisi S, et al. Differentiation of iron deficiency from thalassaemia trait: a new approach. *Haematologica*. 1987;72(5):409-13. PMID:3121463
13. Green R, King R. A new red cell discriminant incorporating volume dispersion for differentiating iron deficiency anemia from thalassaemia minor. *Blood Cells*. 1989;15(3):481-91; discussion 92-5. PMID:2620095
14. d'Onofrio G ZG, Ricerca BM, Mancini S, Mango G. Automated measurement of red blood cell microcytosis and hypochromia in iron deficiency and beta-thalassemia trait. *Archives of Pathology & Laboratory Medicine*. 1992;116(1):84-9. PMID:1734838
15. Sonati Mde F, Grotto HZ, Kimura EM, Costa FF. [Differentiation between heterozygotic beta-thalassemia and iron deficiency anemia]. *Rev Assoc Med Bras*. 1993;39(4):221-3. PMID:8162086
16. Lima CS, Reis AR, Grotto HZ, Saad ST, Costa FF. Comparison of red cell distribution width and a red cell discriminant function incorporating volume dispersion for distinguishing iron deficiency from beta thalassemia trait in patients with microcytosis. *Sao Paulo Med J*. 1996;114(5):1265-9. <https://doi.org/10.1590/S1516-31801996000500005> PMID:9239926
17. Sirdah M, Tarazi I, Al Najjar E, Al Haddad R. Evaluation of the diagnostic reliability of different RBC indices and formulas in the differentiation of the beta-thalassaemia minor from iron deficiency in Palestinian population. *Int J Lab Hematol*. 2008;30(4):324-30. <https://doi.org/10.1111/j.1751-553X.2007.00966.x> PMID:18445163
18. Ehsani MA, Shahgholi E, Rahiminejad MS, Seighali F, Rashidi A. A new index for discrimination between iron deficiency anemia and beta-thalassemia minor: results in 284 patients. *Pak J Biol Sci*. 2009;12(5):473-5. <https://doi.org/10.3923/pjbs.2009.473.475> PMID:19579993
19. Sirachainan N, Iamsirirak P, Charoenkwan P, Kadegasem P, Wongwerawattanakoon P, Sasanakul W, et al. New mathematical formula for differentiating thalassemia trait and iron deficiency anemia in thalassemia prevalent area: a study in healthy school-age children. *Southeast Asian J Trop Med Public Health*. 2014;45(1):174-82. PMID:24964667
20. Miri-Moghaddam E, Sargolzaie N. Cut off Determination of Discrimination Indices in Differential Diagnosis between Iron Deficiency Anemia and beta- Thalassemia Minor. *Int J Hematol Oncol Stem Cell Res*. 2014;8(2):27-32. PMID:24800036 PMCID:PMC4003440
21. Sahli CA, Bibi A, Ouali F, Fredj SH, Dakhlouli B, Othmani R, et al. Red cell indices: differentiation between beta-thalassemia trait and iron deficiency anemia and application to sickle-cell disease and sickle-cell thalassemia. *Clin Chem Lab Med*. 2013;51(11):2115-24. <https://doi.org/10.1515/ccml-2013-0354> PMID:23800659
22. Vehapoglu A, Ozgurhan G, Demir AD, Uzuner S, Nursoy MA, Turkmen S, et al. Hematological indices for differential diagnosis of Beta thalassemia trait and iron deficiency anemia. *Anemia*. 2014;2014:576738. <https://doi.org/10.1155/2014/576738> PMID:24818016 PMCID:PMC4003757
23. Beyan C, Kaptan K, Ifran A. Predictive value of discrimination indices in differential diagnosis of iron deficiency anemia and beta-thalassemia trait. *Eur J Haematol*. 2007;78(6):524-6. <https://doi.org/10.1111/j.1600-0609.2007.00853.x> PMID:17419742
24. Ntaios G, Chatzinkolaou A, Saouli Z, Girtovitis F, Tsapanidou M, Kaiafa G, et al. Discrimination indices as screening tests for beta-thalassemic trait. *Ann Hematol*. 2007;86(7):487-91. <https://doi.org/10.1007/s00277-007-0302-x> PMID:17476506
25. Rathod DA, Kaur A, Patel V, Patel K, Kabrawala R, Patel V, et al. Usefulness of cell counter-based parameters and formulas in detection of beta-thalassemia trait in areas of high prevalence. *Am J Clin Pathol*. 2007;128(4):585-9. <https://doi.org/10.1309/R1YL4B4BT2WCQDGV> PMID:17875509
26. Amendolia SR, Cossu, G., Ganadu, M.L., Golosio, B., Masala, G.L., Mura, G.M. A comparative study of K-Nearest Neighbour, Support Vector Machine and Multi-Layer Perceptron for Thalassemia screening. *Chemometrics and Intelligent Laboratory Systems*. 2003;69:13-20. [https://doi.org/10.1016/S0169-7439\(03\)00094-7](https://doi.org/10.1016/S0169-7439(03)00094-7)
27. Kiss TL, Ali MA, Levine M, Lafferty JD. An algorithm to aid in the investigation of thalassemia trait in multicultural populations. *Arch Pathol Lab Med*. 2000;124(9):1320-3. PMID:10975930
28. Lafferty JD, Crowther MA, Ali MA, Levine M. The evaluation of various mathematical RBC indices and their efficacy in discriminating between thalassaemic and non-thalassaemic microcytosis. *Am J Clin Pathol*. 1996;106(2):201-5. <https://doi.org/10.1093/ajcp/106.2.201> PMID:8712174
29. Sargolzaie N, Miri-Moghaddam E. A local equation for differential diagnosis of beta-thalassemia trait and iron deficiency anemia by logistic regression analysis in Southeast Iran. *Hemoglobin*. 2014;38(5):355-8. <https://doi.org/10.3109/03630269.2014.948187> PMID:25155260
30. Urrechaga E, Aguirre U, Izquierdo S. Multivariable discriminant analysis for the differential diagnosis of microcytic anemia. *Anemia*. 2013;2013:457834. <https://doi.org/10.1155/2013/457834> PMID:24093062 PMCID:PMC3777209
31. Zaghoul A, Al-Bukhari TA, Bajuiafer N, Shalaby M, Al-Pakistani HA, Halawani SH, et al. Introduction of new formulas and evaluation of the previous red blood cell indices and formulas in the differentiation between beta thalassemia trait and iron deficiency anemia in the Makkah region. *Hematology*. 2016;21(6):351-8. <https://doi.org/10.1080/10245332.2015.1133753> PMID:26907523
32. Bordbar E, Taghipour M, Zucconi BE. Reliability of Different RBC Indices and Formulas in Discriminating between beta-Thalassemia Minor and other Microcytic Hypochromic Cases. *Mediterr J Hematol Infect Dis*. 2015;7(1):e2015022. <https://doi.org/10.4084/mjhid.2015.022> PMID:25745549 PMCID:PMC4344165
33. Hisham A.Getta HAY, Hameed M. Said Hi & Ha, are new indices in differentiation between iron deficiency anemia and beta-Thalassaemia trait /A Study in Sulaimani City-Kurdistan/Iraq *Journal of Dental and Medical Sciences* 2015;14(7):67-72.
34. Hoffmann JJ, Urrechaga E, Aguirre U. Discriminant indices for distinguishing thalassemia and iron deficiency in patients with microcytic anemia: a meta-analysis. *Clin Chem Lab Med*. 2015;53(12):1883-94. <https://doi.org/10.1515/ccml-2015-0179> PMID:26536581
35. Schoorl M, Schoorl M, van Pelt J, Bartels PC. Application of Innovative Hemocytometric Parameters and Algorithms for Improvement of Microcytic Anemia Discrimination. *Hematol Rep*. 2015;7(2):5843. <https://doi.org/10.4081/hr.2015.5843> PMID:26331001 PMCID:PMC4508552
36. Burges C. Data mining and knowledge discovery. A tutorial on support vector machines for pattern recognition. 1998:121-67.
37. Angastiniotis M, Eleftheriou A, Galanello R, Hartevelde CL, Petrou M, Traeger-Synodinos J, et al. In: Old J, ed. *Prevention of Thalassaemias and Other Haemoglobin Disorders: Volume 1: Principles*. 2nd ed. Nicosia, Cyprus; 2013.
38. Rund D, Filon D, Strauss N, Rachmilewitz E, Oppenheim A. Mean corpuscular volume of heterozygotes for beta-thalassemia correlates with the severity of mutations. *Blood*. 1992;79(1):238-43. PMID:1728311
39. Weatherall DJ, Clegg JB. Inherited haemoglobin disorders: an increasing global health problem. *Bull World Health Organ*. 2001;79(8):704-12. PMID:11545326 PMCID:PMC2566499
40. Wasi P, Disthasongchan P, Na-Nakorn S. The effect of iron deficiency on the levels of hemoglobins A2 and E. *J Lab Clin Med*. 1968;71(1):85-91. PMID:5635011
41. Zaghoul A, Al-Bukhari TA, Bajuiafer N, Shalaby M, Al-Pakistani HA, Halawani SH, et al. Introduction of new formulas and evaluation of the previous red blood cell indices and formulas in the

- differentiation between beta thalassemia trait and iron deficiency anemia in the Makkah region. *Hematology*. 2016.  
<https://doi.org/10.1080/10245332.2015.1133753> PMID:26907523
42. Rahim F, Keikhaei B. Better differential diagnosis of iron deficiency anemia from beta-thalassemia trait. *Turk J Haematol*. 2009;26(3):138-45. PMID:27265497
43. Romero Artaza J, Carbia CD, Ceballo MF, Diaz NB. [Red cell distribution width (RDW): its use in the characterization of microcytic and hypochromic anemias]. *Medicina (B Aires)*. 1999;59(1):17-22.
44. Ismail M, Patel Nisar G. Evaluation of the Diagnostic Accuracy of Twelve Discrimination Indices for Differentiating  $\beta$ -thalassemia Trait from Iron Deficiency Anemia. *Indian Journal of Public Health Research & Development*. 2016;7(1):104.  
<https://doi.org/10.5958/0976-5506.2016.00021.8>