ORIGINAL ARTICLE



Prevalence of β -haemoglobinopathies in Eastern India and development of a novel formula for carrier detection

Sandeep Saha¹ · Shouriyo Ghosh¹ · Keya Basu² · Maitreyee Bhattacharyya¹

Received: 23 April 2020 / Accepted: 1 July 2020 / Published online: 8 July 2020 © Springer-Verlag GmbH Germany, part of Springer Nature 2020

Abstract

Beta thalassemia is the commonest inherited monogenic disorder and India forms a major part of the thalassemia belt in the world. Effective screening programmes and proper prenatal diagnosis are the only ways to eradicate the disease. Haemoglobin high-pressure liquid chromatography (HPLC) is the screening method employed for detection of thalassemia carriers. The aim of this paper was to investigate the prevalence and demographic profiles of thalassemia carriers in Eastern India. Haemogram data was used to develop a novel haematological index for thalassemia carrier detection. This was a retrospective analysis of data collected from 21,695 individuals at thalassemia screening camps. The prevalence of the common hemoglobinopathies and their regional variations were noted. Using stepwise discriminant analysis, RBC indices were analysed to develop a novel formula to help distinguish β -thalassemia carriers from normal individuals. Out of the total population analysed, β -hemoglobinopathy carrier prevalence was around 10%. HbE was the commonest hemoglobinopathy in the region. There was no difference in the prevalence rates of thalassemia carriers amongst religious communities. The formula generated was expressed as outcome = RBCx6.59 + MCHx0.527 - (HCTx0.782 + MCHCx0.395 + RDWx0.02 + 1.365). Any integer value more than 0.5 indicated the probability of carrier status while a value less than 0.5 was normal. The formula had a negative predictive value of >96% and an accuracy of 89%. This study provides an audit of the common β -hemoglobinopathies in Eastern India. The regional variations and prevalence of hemoglobinopathies highlighted here may guide public health efforts. The formula generated from RBC indices can be used to screen individuals prior to HPLC testing, thereby reducing overall costs of thalassemia screening in the country.

Keywords Thalassemia screening · Demographic data · Eastern India · Novel index · Haemogram parameters

Introduction

Beta thalassemia, the most common inherited monogenic disorder, is a major public health issue in the tropics and is being increasingly recognized worldwide. [1, 2] Considerable expenditure of public money is needed for prevention, control and treatment of the disease. [3, 4] India has the largest number of thalassemia major children worldwide, nearly 150,000, with addition of 10,000–15,000 new cases every year. In addition, there are around 42 million β-thalassemia carriers in the country with an average prevalence rate of 3-4%. [5] Although certain haemoglobin (Hb) variants are predominantly found in particular areas of the country (for, e.g. HbE and HbE-beta thalassemia is mainly found in West Bengal and north-eastern parts of the country), due to population migration and admixture, the entire spectrum of thalassemia and Hb variants are now found all over the country. [6] In an inherited disorder with variable genotype-phenotype associations, the only way to eradicate the disease is to have a robust screening program in place, one which can conclusively detect carrier state and help genetic counselling of a couple planning their family. Thus, carrier detection has a major impact in the control of this disease. The current population screening programs are based on Hb high-pressure liquid chromatography (HPLC) of haemoglobin samples. [7] The aim of this paper was to analyse the data collected at thalassemia screening camps and determine carrier prevalence rates, the demographic profiles and the distribution of carrier state across various communities in the region. RBC indices were analysed to

Shouriyo Ghosh shouriyo.ghosh@gmail.com

¹ Institute of Haematology and Transfusion Medicine, Medical College Hospital, 88, College Street, Kolkata, West Bengal 700073, India

² Calcutta National Medical College, Kolkata, West Bengal, India

generate a novel formula to distinguish between normal individuals and β -thalassemia carriers.

Material and methods

Study design This was a retrospective data analysis of 21,695 individuals screened for thalassemia. Duration of the study was over a period of 5 years (January 2014-December 2019). Informed consent was taken from all participants. Demographic data, HPLC data and haemogram parameters were noted for analysis. The individuals in the study population were screened at camps, outreach centres, schools and antenatal clinics organised by the Thalassemia Control Unit of the Institute of Haematology and Transfusion Medicine (IHTM), Medical College Hospital, Kolkata, West Bengal. IHTM, which is the nodal centre for thalassemia control in the state, performs routine antenatal screening for thalassemia in the form of pre-conceptional testing, prenatal testing at 8-12 weeks of gestation and chorionic villus sampling (CVS) in applicable cases. However, this study was done to determine the prevalence of thalassemia carriers and generate a novel screening formula only; thus, prenatal diagnosis and CVS data have not been included here. Thalassemia major patients detected by HPLC were also excluded.

Haemogram and HPLC analysis Blood samples were run on the Sysmex KX 21 haematology analyser [Sysmex Corporation, Japan] and HPLC was done on the Bio-Rad Variant Haemoglobin Testing System (Beta-thalassemia short program) [Bio-Rad, California, USA]. Peripheral blood smear examination was done for validating the haemogram results. For each of the RBC parameters, viz. haemoglobin, mean corpuscular volume (MCV), mean corpuscular haemoglobin (MCH), mean corpuscular haemoglobin concentration (MCHC), RBC count and red-cell distribution width (RDW), mean values with ± 2 standard deviations were recorded. Two groups were made for the subsequent analysis, viz. group A with normal Hb HPLC and group B with β hemoglobinopathy carrier by Hb HPLC.

Statistical analysis The data was collected, compiled and analysed using SPSS version 19.0 [IBM Inc., Armonk, New York, USA]. The qualitative variables were expressed in terms of percentages. The quantitative variables were both categorised and expressed in terms of percentages or in terms of mean and standard deviations. Difference between two proportions was analysed using Chi square or fisher exact test. All analysis was two-tailed and the significance level was set at 0.05. Stepwise discriminant analysis was applied to RBC indices to generate a formula to distinguish between thalassemia carriers and normal individuals.

Results

Demographic data A total of 17,764 female and 3931 male individuals were screened. Mean age of the population was 22 years (age range: -1 to 75 years). Most of the respondents in the study were antenatal mothers (n = 14,891; 68.6%) of the total number screened). Out of the total of 21,695 screened individuals, 19,432 individuals had normal Hb HPLC and 2263 were β-thalassemia/hemoglobinopathy carriers (by HPLC) and comprised beta thalassemia trait (941/2263; 41.5%), HbE trait (1128/2263; 49.84%), HbS trait (100/2263; 4.41%), HbD trait (18/2263; 0.79%) and delta beta trait (18/2263; 0.79%). Other rare types constituted 58/2263 (2.56%) cases and included Hb Lepore and HPFH trait. HbE trait and beta thalassemia trait comprised 5.2% and 4.3% of the total individuals, respectively, making HbE trait the most common hemoglobinopathy prevalent in this region. Combining the above two most common traits, the average thalassemia carrier rate in the screened population was around 10%. Carrier rate of all hemoglobinopathies in females was 9.2% while in males, it was 15.9%. The major districts covered by the camps were those in the vicinity of the hospital and included Kolkata (11,017/21,695-50.8%) and South-24-Parganas (9160/21,695-42.2%). The other districts, viz. North-24-Parganas, Coochbehar, Birbhum, Bankura, Hooghly and Darjeeling to name a few, comprised the remaining 7% of the total population. There was no significant difference in the prevalence of carrier state in the two major religious communities (Hindus and Muslims) in the region, and they comprised 47.68% (n = 1079/2263) and 52.01% (n = 1177/2263) of the total carrier population, respectively (Tables 1 and 2).

Analysis of haemogram parameters The means of all the haemogram parameters were compared between the normal individuals and thalassemia carriers. The difference of means was found to be statistically significant for all parameters (p < 0.001) (Table 3). As expected, the mean haemoglobin, MCV, MCH, MCHC and haematocrit were higher in normal individuals while the RDW and RBC counts were higher in the carriers (Table 2).

Development of the novel index After the analysis of haemogram parameters, stepwise discriminant analysis was applied to the RBC indices. The outcome was then expressed as an equation:

scale80%Outcome

- = RBCx6.59 + MCHx0.527
- -(HCTx0.782 + MCHCx0.395 + RDWx0.02 + 1.365)

Parameters		Normal <i>N</i> = 19,43 (%)	Normal $N = 19,432$ B-Thalassemia trait (%) $N = 941$ (%)	HbE trait $N = 112$ (%)	8 δß trait N=18(%)	HbE trait $N = 1128$ 58 trait $N = 18$ Hb S trait $N = 100$ Hb D trait $N = 18$ Other rare types (%) (%) (%) N = 58 (%)) Hb D trait <i>N</i> = 18 (%)	S Other rare types $N = 58 (\%)$	Total
Age (years) mean ± 2 SD	an ± 2 SD	22.09 ± 5.65	24.22 ± 7.93	22.21 ± 9.07	20.83 ± 11.78	22.43 ± 9.12	21.28 ± 12.34	20.5 ± 9.47	22.18 ± 6.06
Sex	Male	3305 (15.2)	259 (1.19)	311 (1.43)	4 (0.01)	20 (0.09)	6 (0.02)	26 (0.11)	3931 (18.11)
	Female	16,127 (74.3)	682 (3.14)	817 (3.76)	14 (0.06)	80 (0.36)	12 (0.05)	32 (0.14)	17,764 (81.88)
District	Kolkata	10,308 (47.5)	268 (1.23)	350 (1.61)	7 (0.03)	52 (0.02)	9 (0.04)	23 (0.1)	11,017 (50.7)
	South-24-P	7787 (35.8)	602 (2.77)	682 (3.14)	9 (0.04)	42 (0.19)	8 (0.03)	30 (0.13)	9160 (42.22)
	North-24-P	635 (2.92)	23 (0.10)	43 (0.19)	0 (0)	4 (0.01)	1 (0.004)	1 (0.004)	707 (3.25)
	Others	702 (3.23)	48 (0.22)	53 (0.24)	2 (0.09)	2 (0.09)	0 (0)	4 (0.01)	811(3.73)
Religion	Hindu	8990 (41.4)	502 (2.31)	498 (2.29)	12 (0.05)	28 (0.12)	12 (0.05)	27 (0.12)	10,069 (46.41)
	Muslim	10,356 (47.7)	436 (2)	627 (2.89)	5 (0.023)	72 (0.33)	6 (0.02)	31 (0.14)	11,533 (53.15)
	Christian	73 (0.3)	2 (0.09)	2 (0.09)	0 (0)	0 (0)	0 (0)	0 (0)	77 (0.35)
	Sikh	9 (0.04)	1 (0.004)	0 (0)	1 (0.004)	0 (0)	0 (0)	0 (0)	11 (0.05)
	Jain	1 (0.004)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (0.004)
	Buddhist	3 (0.01)	0 (0)	1 (0.004)	0 (0)	0 (0)	0 (0)	0 (0)	4 (0.01)
Respondent	Antenatal mother	13,756 (63.4)	496 (2.28)	559 (2.57)	8 (0.03)	50 (0.23)	6 (0.02)	16 (0.07)	14,891 (68.63)
Category	Children	128 (0.58)	0 (0)	7 (0.03)	2 (0.09)	0 (0)	0 (0)	3 (0.01)	140 (0.64)
	(age < 18 years)								
	Family member of thalassemic	1019 (4.69)	144 (0.66)	250 (0.11)	4 (0.01)	27 (0.12)	3 (0.01)	9 (0.04)	1456 (6.71)
	Family member of carrier	14 (0.06)	1 (0.004)	1 (0.004)	0 (0)	1 (0.004)	0 (0)	(0) 0	17 (0.07)
	Premarital group	2948 (13.58)	48 (0.22)	109 (0.5)	0 (0)	5 (0.023)	6 (0.02)	6 (0.02)	3122 (14.39)
	Post-marital group	231 (1.06)	15 (0.06)	7 (0.03)	0 (0)	0 (0)	1 (0.004)	0 (0)	254 (1.17)
	Others	1336	237	195	4	17	2	24	1815

 Table 1
 Demographic details of the study population

Percentage displayed is of the total study population (n = 21, 695)

 Table 2
 Hb and RBC indices of the study population

Parameters	Normal individuals $N = 19,432$	β-Thalassemia trait N = 941	HbE trait $N = 1128$	δß trait $N = 18$	Hb S trait $N = 100$	Hb D trait $N = 18$	Other rare types $N = 58$	Total
Hb (g/dl) mean ± 2 SD	11.65 ± 1.59	10.48 ± 1.63	11.44 ± 1.59	11.04 ± 1.49	11.60 ± 1.57	12.39 ± 1.61	10.67 ± 1.99	11.59 ± 1.61
RBC count (mill/cumm) mean ± 2 SD	4.1 ± 0.58	4.87 ± 0.77	4.56 ± 0.68	4.79 ± 0.71	4.32 ± 0.66	4.43 ± 0.61	4.37 ± 0.65	4.16 ± 0.62
MCV (fl) mean ± 2 SD	86.32 ± 5.93	69.11 ± 3.96	76.86 ± 4.83	72.34 ± 3.48	81.03 ± 7.78	83.54 ± 5.91	75.6 ± 7.71	85.01 ± 7.09
$\begin{array}{c} \text{MCH (pg) mean} \\ \pm 2 \text{ SD} \end{array}$	28.53 ± 2.48	21.58 ± 1.46	25.2 ± 1.85	23.17 ± 1.57	27.09 ± 3.23	28.07 ± 2.37	24.46 ± 3.07	28.03 ± 2.89
MCHC (g/dl) mean ± 2 SD	33.02 ± 1.22	31.23 ± 0.83	32.78 ± 0.94	32.01 ± 1.27	33.36 ± 1.19	33.58 ± 0.82	32.3 ± 1.26	32.93 ± 1.25
RDW-CV (%) mean ± 2 SD	14.74 ± 1.77	16.89 ± 1.95	15.06 ± 1.55	19.05 ± 3.53	14.7 ± 1.58	15.2 ± 1.66	17.84 ± 3.22	14.86 ± 1.84

An integer value more than 0.5 meant the probability of the individual being a thalassemia carrier while value of 0.5 or less implied a normal individual. The formula was validated on a validation dataset of 1200 individuals, also collected through screening camps and found to have a sensitivity 74.51%, specificity of 90.82%, positive predictive value of 47.99%, negative predictive value of 96.91%, and an accuracy of 89.2%.

Discussion

There are some studies on the prevalence and distribution of thalassemia traits in India and the numbers depicted closely match those of ours. Sachdev et al. found a beta-thalassemia trait prevalence rate of 8.9% in northern India while Dolai et al. found a prevalence rate of 10.38% and 4.30% of β -thalassemia trait and HbE trait, respectively. [8, 9] β -thalassemia trait appears to be the most common hemoglobin-opathy, while in our study, the prevalence of HbE trait was found to be slightly higher as compared to β -thalassemia trait. [8–10] This is in-sync with the increased regional prevalence

of HbE in Eastern India. There were no community specific or regional variations in the prevalence of thalassemia carriers in the study population. Consanguinity is prevalent in parts of the country where this study was done and becomes an important issue while evaluating a couple for prenatal/preconceptional testing. Since prenatal diagnosis was not the primary focus in this study, we did not evaluate this aspect. Contrary to the principles of thalassemia genetics and inheritance patterns [11], a slightly higher carrier frequency was noted in males (15.9% versus 9.2% in females) and is due to the fact that they were husbands of the women detected to be thalassemia carriers at antenatal clinics.

In this study, we used discriminant analysis similar to Zaghloul et al. and Schoorl et al. and generated the formula utilizing most of the RBC indices. [12, 13] The formula has a very good specificity, negative predictive value and accuracy and can differentiate true carriers from normal individuals in the general population. It was generated from a large dataset, representing all age groups, sexes, haemoglobin levels and ethnicities. It is a more inclusive formula, taking into consideration a large spectrum of haemogram values and a wide variety of thalassemia carrier states. Previously, numerous attempts have been made to distinguish normal individuals,

Parameter	Normal individuals (group A)		Thalassemia traits (group B)		P value
	Mean	±2 SD	Mean	± 2 SD	
Haemoglobin (g/dl)	11.65	1.59	11.03	1.68	< 0.001
RBC count (mill/cumm)	4.10	0.58	4.67	0.74	< 0.001
HCT (%)	35.27	4.53	34.34	4.99	< 0.001
MCV (fl)	86.32	5.93	73.81	6.30	< 0.001
MCH (pg)	28.53	2.48	23.77	2.65	< 0.001
MCHC (g/dl)	33.02	1.22	32.15	1.21	< 0.001
RDW-CV (%)	14.74	1.77	15.91	2.06	< 0.001

Table 3Analysis of haemogramparameters in the two groups

individuals with iron deficiency and thalassemia carriers based on RBC parameters. Of the available formulae/indices, Shine and Lal et al. provided an index which utilized MCV and MCH to discriminate carriers from normal individuals. Their index gave a sensitivity of 98.74% and a specificity or accuracy of 88%. [14] HbA2 \geq 4% is the usual cut-off used for detection of β -thalassemia trait while values ranging between 3.6 and 3.9% are considered borderline and the individuals with these values are advised partner screening as part of pre-natal counselling. It is also known that the HbA2 levels are lower in individuals with concomitant iron deficiency and it is recommended that IDA be treated before performing HPLC. [15] However, Khera et al. in 2015 reported that, although HbA2 values could be low in iron deficiency anaemia, they were discriminatory enough to clearly identify true carriers. [16] Roth et al., in 2018, published an excellent review of all the available literature on this topic and generated a formula using a unique algorithm called support vector machine (SVM). Sensitivity using the SVM approach was > 99%with an accuracy of 88%. They used only MCV and MCH in the above-mentioned mathematical model to derive results. It is interesting to note that the study analysed data after excluding cases with haemoglobin levels less than 9 g/dl, as a lower Hb value was thought to interfere with accuracy of Hb HPLC. [17] Das et al. have recently published a decision tree to distinguish β-thalassemia trait and HbE from normal individuals using scores developed by artificial neural networks. They used all the RBC indices and achieved very good sensitivities and negative predictive values of 100% with false positive rates ranging from 20 to 40% in validation cohorts. [18] We also used all the RBC indices and wide range of haemoglobin levels (7 g/dl to 15 g/dl) to obtain satisfactory results. Our dataset also included age groups in the range of 1-75 years and is therefore more inclusive and considers the age variations of RBC indices.

This study has few limitations. A major one is the lack of iron profile data of the individuals screened. This precluded the stratification of the patients into iron deficient and iron replete groups to allow a more meaningful evaluation of RBC indices and generation of the formula. Another limitation is the skewness of the respondents towards predominantly antenatal mothers and the apparent increase in the prevalence of male carriers due to reasons discussed above. In addition, the stepwise discriminant analysis did not give results comparable to methods such as artificial neural networks used by other authors to achieve the same goal.

Conclusion

Having knowledge of the demographics of thalassemia carrier state in the regions where the control projects are running is of utmost importance and help in identifying endemic zones, changes in prevalence, common variants, effectiveness of the screening and control programme and above all help in formulating health-care, administrative and clinical practice guidelines. In addition, specialized tests like pre-natal screening for mutations can be designed in a much more costeffective way, to specifically pick up the common mutations or haemoglobin variants in that region. For example, regional differences exist in India and Eastern India with a higher prevalence of HbE than the national average is a case in point. The only method of eradicating thalassemia is by proper screening and genetic counselling of couples. The future of screening for thalassemia would be to develop a robust computer algorithm (using machine learning and artificial intelligence) which would utilize RBC parameters to distinguish thalassemia carriers from normal individuals with considerable accuracy. Such software could be incorporated within haematology auto-analysers, mobile phone and web-based applications etc. It would be much cheaper than Hb HPLC, and HPLC could then be used only for confirmation of carrier status in the flagged individuals.

Compliance with ethical standards

This study was a retrospective data analysis and did not involve animal experimentation in any form. Informed consent was duly taken from all participants and the study was approved by the Institutional Ethical Committee.

Conflict of interest The authors declare that they have no conflict of interest.

References

- 1. Williams TN, Weatherall DJ (2012) Health burden of the hemoglobinopathies. Cold Spring Harb Perspect Med 2((9):a0116):1–14
- Kantharaj A, Chandrashekar S (2018) Coping with the burden of thalassemia: aiming for a thalassemia free world. Glob J Transfus Med 3(1):1
- Colah R, Italia K, Gorakshakar A (2017) Burden of thalassemia in India: the road map for control. Pediatr Hematol Oncol J 2(4):79– 84
- Italia KY, Jijina FJ, Merchant R, Panjwani S, Nadkarni AH, Sawant PM, Nair SB, Ghosh K, Colah RB (2009) Response to hydroxyurea in β thalassemia major and intermedia: experience in western India. Clin Chim Acta 407(1–2):10–15
- MOHFW (2016) National Health Mission Guidelines on Hemoglobinopathies in India: Prevention and control of hemoglobinopathies in india. [Internet]. Available from: https://nhm.gov.in/ images/pdf/programmes/RBSK/Resource_Documents/ Guidelines on Hemoglobinopathies in India.pdf
- Ghosh K, Colah R, Manglani M, Choudhry VP, Verma I, Madan N et al (2014) Guidelines for screening, diagnosis and management of hemoglobinopathies. Indian J Hum Genet 20(2):101–119
- Waters HM, Howarth JE, Hyde K, Goldstone S, Cinkotai KI, Kadkhodaei-Elyaderani MRJ (1998) An evaluation of the Bio-Rad Variant Haemoglobin Testing System for the detection of haemoglobinopathies. Clin Lab Haematol 20(1):31–40

- Sachdev R, Dam AR, Tyagi G (2010) Detection of Hb variants and hemoglobinopathies in Indian population using HPLC: report of 2600 cases. Indian J Pathol Microbiol 53(1):1–6
- Dolai TK, Dutta S, Bhattacharyya M (2012) Original article prevalence of hemoglobinopathies in rural Bengal, India. Haemoglobin;36(March 2011):57–63
- Rao S, Kar R, Gupta SK, Chopra A, Saxena R (2010) Spectrum of haemoglobinopathies diagnosed by cation exchange- HPLC & modulating effects of nutritional deficiency anaemias from North India. Indian J Med Res 132:513–519
- 11. Cao A, Galanello R (2010) Beta-thalassemia. Genet Med 12:61–76 Nature Publishing Group
- Schoorl M, Schoorl M, Van Pelt J, Bartels PCM (2015) Application of innovative hemocytometric parameters and algorithms for improvement of microcytic anemia discrimination. Hematol Rep 7(2): 52–55
- Zaghloul A, Al-bukhari TAMA, Bajuaifer N, Shalaby M, Al-Pakistani HA, Halawani SH et al (2016) 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. 21(6): 351–358

- Shine IAN (1977) A strategy to detect beta-thalassaemia minor. Lancet. 1(8013):692–694
- Wasi P, Disthasongchan P, Na-Nakorn S (1968) The effect of iron deficiency on the levels of hemoglobins A2 and E. J Lab Clin Med 71(1):85–91
- Khera R, Singh T, Khuana N, DA Gupta N (2015) HPLC in characterization of hemoglobin profile in thalassemia syndromes and hemoglobinopathies: a clinicohematological correlation. Indian J Hematol Blood Transfus 31(1):110–115
- Roth IL, Lachover B, Koren G, Levin C, Zalman L, Koren A (2018) Detection of β-thalassemia carriers by red cell parameters obtained from automatic counters using mathematical formulas. Mediterr J Hematol Infect Dis 10(1):1–10
- Das R, Datta S, Kaviraj A, Sanyal SN, Nielsen P, Nielsen I et al (2020) A decision support scheme for beta thalassemia and HbE carrier screening. J Adv Res 24:183–190. [Internet]. Available from. https://doi.org/10.1016/j.jare.2020.04.005

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.