

Risk Score Estimation of Diabetic Retinopathy: Statistical Alternatives using Multiple Logistic Regression

Vaitheeswaran Kulothungan^{1,2*}, Ramakrishnan R³, Subbiah M⁴ and Rajiv Raman⁵

¹Research Scholar, Manonmaniam Sundaranar University, Tirunelveli 627 012, Tamil Nadu, India

²National Center for Disease Informatics and Research (NCDIR), Bangalore 562 110, Karnataka, India

³National Institute of Epidemiology (ICMR), Ayapakkam, Chennai 600 077, Tamil Nadu, India

⁴Department of Mathematics, L.N Government College, Ponneri 601 204, Tamil Nadu, India

⁵Shri Bhagwan Mahavir Vitreoretinal Services, Sankara Nethralaya, 18, College Road, Chennai 600 006, Tamil Nadu, India

Abstract

People with type II diabetes are having more chances to develop diabetic retinopathy which is generally viewed as a multi-factorial disease. Identifying the risk of any disease, is very important for health care planning and creating score cards for identifying the risk of any disease is pervasive in medical diagnostics. This involves statistical techniques using parameter estimation of multivariable models such as linear regression, logistic regression or Cox proportional hazards regression. Geographic and/or disease specific methods for risk score estimation provide a scope to develop and evaluate new possible statistical methods for risk score analysis. This work explores the weighted scoring procedure through logistic regressions to develop two methods using Wald statistic and maximum regression coefficient by precluding the selection of protective risk factors. Further, to avoid the numerical errors due to interim rounding of digits in any computations the study includes a standard method that avoids such rounding of digits. Three widely applicable methods for score estimation of different diseases have also been considered for comparative study. All these methods are applied to Sankara Nethralaya Diabetic Retinopathy Epidemiology and Molecular Genetics Study III, a cross-sectional study to estimate the prevalence and risk factors for diabetic retinopathy in rural south India and then validated by comparing with methods used in Australian Type 2 Diabetes Risk Assessment Tool. Results have indicated that the new methods are more suitable in score estimation for diabetic retinopathy by considering the statistical property of the methods.

Keywords: Diabetic retinopathy; Logistic regression; Risk score; Wald statistic; Weighted scores

Introduction

Standardized risk score systems are used to create a system which aid in early diagnosis of a disease. Classifying individuals at risk for any disease would promote the efficiency of health system. This is relevant in Diabetes, as with these scores one could accurately predict complications and prevent their progression [1]. The epidemic of Diabetes Mellitus (DM), in particular type 2 diabetes mellitus, is assuming significant proportion in developing countries, such as India [2,3]. International Diabetes Federation (IDF) has projected that number of people with diabetes in India would rise from 65.1 million in 2013 to 109 million in 2035 [4].

Diabetic retinopathy (DR) is the most common ocular complication of diabetes. Blindness due to retinopathy is the major disability in patients with diabetes [5]. DR is often asymptomatic until a significant structural and irreversible damage occurs. Late diagnosis of DR results in significant socio-economic burden to the patient [6]. Several studies have shown that early and regular fundus examinations are important in screening, diagnosing, and monitoring DR [5-7]. Recently, risk scores for diabetes retinopathy based on anthropometric, demographic and clinical variables have been suggested to screen for DR [8,9]. However, a common risk score cannot be applied for all populations due to ethnicity differences.

Many alternative methodologies have been tried to estimate the risk score by assigning the different types of weighing procedure through logistic regression [10,11]. These scoring systems are developed using the parameters which are found to be significant at 20% level through the multiple logistic regression using stepwise backward elimination. Subsequently, the score systems are scaled in the different adoptable measures such as by multiplying the regression coefficients by 10 and

rounding to the nearest integer [12,13]. By doing so, the constant corresponded to one point in the risk score system.

For each risk factor, its distance from the base category in regression coefficient units is divided by this constant and rounded to the nearest integer to get its point value [13]. Then, by dividing the coefficient for each variable in the final model by the lowest coefficient, then multiplying by 2 (all factors are significant) and rounding to whole number, a component is obtained [14]. Similarly, by dividing coefficients by the absolute value of the smallest coefficient in the model and rounding up to the nearest integer, another component is obtained. The overall risk score is calculated by adding each component together from the half sum of the two smallest coefficients in the model [15,16].

With these collective observations among the available methods, this study aims to identify suitable alternatives using statistically acceptable modifications. The objectives are (1) to address the errors due to interim rounding of digits and its impact on estimates between different weight methods; (2) to develop and validate newer methods using Wald statistic and maximum regression β -coefficient that can

***Corresponding author:** Vaitheeswaran K, M.Sc., M.Phil., Scientist 'C' (Statistics), National Center for Disease Informatics and Research (NCDIR), Indian Council of Medical Research (ICMR), Kannamangala Post, Bangalore 562 110, Karnataka, India, Tel: +91-08-28467643; Fax: +91-08-28467644; E-mail: vaitheeswaran81@gmail.com

Received November 07, 2014; **Accepted** December 23, 2014; **Published** December 31, 2014

Citation: Vaitheeswaran K, Ramakrishnan R, Subbiah M, Raman R (2014) Risk Score Estimation of Diabetic Retinopathy: Statistical Alternatives using Multiple Logistic Regression. J Biomet Biostat 5: 211. doi:10.472/2155-6180.1000211

Copyright: © 2014 Vaitheeswaran K, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

easily be applied to identify the risk of DR for an individual of any diversity and (3) to compare the performance of different methods through exploring different weightier procedure in developing risk score systems.

The presentation of this article is as follows: various methods for risk estimating are listed in Section 2; details of the data used in the study and the statistical analysis have been presented in Section 3 and the concluding remarks are in Section 4.

Methods for Risk Score Estimation

The general approach for developing a risk score system is explained in the context of multiple logistic regressions that has the form

$$y_i = X_i' \beta + \varepsilon_i$$

Where the response variable y_i is a Bernoulli random variable that takes on the values either 0 or 1; $X_i' = [1, X_{i1}, X_{i2}, \dots, X_{ip}]$ is the observed quantities of the risk factors $X_1, X_2, \dots, X_p$ that can be continuous or indicator/dummy variables reflecting dichotomous risk factors or categories of risk factors; and $\beta' = (\beta_0, \beta_1, \beta_2, \dots, \beta_p)$ are the estimates of the regression coefficients based on the regression model; the errors (ε_i) has mean zero, non constant variance and need not follow normal distribution.

In order to determine points, the continuous factors (listed in X) are converted into categories based on research interest. Then, we determine the reasonable categories for each risk factor to serve as the reference category. The reference category for each risk factor is assigned as zero in the scoring system.

Scoring systems are using different types of weights W_i calculate the scores using logistic regression method, where the constant W_i is fixed to calculate the points in the score system along with β_i coefficients. Table 1 shows the description of some of the existing methods that are widely applied in various medical risk estimations. Standard method (SM) provides the risk scores without any interim round off for decimal places.

The methods M_1 - M_3 considered the β coefficient that is statistically significant in logistic modeling. The two naive approaches are M_1 and M_3 . However, Wald statistic provides more information regarding the significance among β coefficients. Hence this work identifies a method (M_2) based on Wald statistic; that is by choosing the β coefficient of risk factor correspond to highest value of Wald statistic in the output of logistic regression say β_j .

Thus, W_i for method M_2 in the range $(-\infty, \infty)$ will be

$$W_i = \left[100 \left(\frac{\beta_i}{\beta_j} \right) \right]$$

Method No	Extracted from	W_i	Range in set of integers
M_1	Glumer [11]	$[10\beta_i]$	$(-\infty, \infty)$
M_2	Lei Chen [14]	$\left[2 \left(\frac{\beta_i}{\text{Min}\{\beta_i\}} \right) \right]$	$(-\infty, \infty) - \{0\}$
M_3	Sugioka [15]	$\left[2 \left(\frac{\beta_i}{\text{Mean of smallest two } \beta\text{'s}} \right) \right]$	$(-\infty, \infty) - \{0\}$

Table 1: Description of existing methods which are widely applied in medical risk estimations.

Also compared to M_2 , this study has chosen max β 's instead of min $|\beta|$'s and then multiply by 100; in the M_4 and M_5 , the β_j will precludes the selection of protective factors (i.e., negative regression coefficients). Hence W_i for method M_5 is

$$W_i = \left[100 \left(\frac{\beta_i}{\text{Max}\{\beta_i\}} \right) \right]$$

with W_i ranges over the integers from $-\infty$ to 100; In all these methods $[.]$ indicates the nearest integer function; accordingly $[1.25]=1$, $[1.50]=2$ and so on.

Subsequently, probability of the risk is computed for each factor to determine the risk of the individuals (or probability of developing an event) and is calculated as

$$\hat{p} = \frac{1}{1 + \exp(-\sum_{i=0}^p \beta_i X_i)}$$

The basic idea of the point system is to approximate the contribution of the risk factors in the estimate of risk, specifically to estimate $\sum_{i=0}^p \beta_i X_i$

which is the component of each model shown above, which depends on the specific risk factor profile under consideration. The risk estimates in the points system associated with specific risk factor profiles are computed by substituting the product of the total number of points and the constant, W_i , which approximates $\sum_{i=0}^p \beta_i X_i$ into the appropriate formula (e.g. logistic regression equation) to estimate the risk.

The proposed risk score model algorithms tested for model goodness-of-fit which is evaluated using the Hosmer-Lemeshow test statistic [17] and overall predictive accuracy of the model is assessed using the c -statistic, which has similar value or equivalent to the area under the receiver operating characteristic (ROC) curve [18]. Bland and Altman test is performed to evaluate agreement between estimated risk in rounding and non-rounding methods [19]. In additional, intra class correlation coefficient (Cronbach's alpha) is also used to assess the intersession variability [20,21]. A p value of <0.05 is considered statistical significant. All these evaluations have been implemented through the statistical software (SPSS for Windows, ver.14.0 SPSS Inc, Chicago, IL, USA).

Results

Sankara Nethralaya-Diabetic Retinopathy Epidemiology and Molecular Genetic Study (SN-DREAMS III) is a population-based, cross-sectional, study to estimate the prevalence and risk factors of diabetes and diabetic retinopathy in the South-Indian population. The detail methodology and study design of SN-DREAMS III is given elsewhere [22]. The study population is selected by multi-stage cluster sampling procedure where each cluster are having of 1,200-2,000 population, who are selected with probability proportional to size (PPS) and the sampling weightage (reciprocal of sampling fraction) is considered into these methodologies. The diabetic retinopathy (DR) data is considered as a binary response variable (0=No DR, 1=DR) with the 12 independent variables included in the model. From the study, 1329 subjects with diabetes are included in the present study.

The risk factors have further grouped into demographic, Anthropometric measurements and biochemical factors. The demographic risk factors studied are age, gender and physical activity. Most of the studies use four groups to break the continuous variable,

age; for example, 40-49, 50 to 59 years, 60 to 69 years, and 70 years. But this study has considered two categories less than 55 as younger and more than 55 as elder group. The younger age group is used as the base category. The systemic risk factors studied included the duration of diabetes mellitus, user of insulin, family history of diabetes mellitus and history of hypertension.

Anthropometric measurements, including weight and height are obtained using standardized techniques and the body mass index (BMI) is calculated using the formula: weight (kg)/height (m²). Based on the BMI, individuals are classified as lean (male<20, female<19), normal (male 20-25, female 19-24), overweight (male 25-30, female 24-29) or obese (male>30, female>29). Glycemic control is categorized as normal (HbA_{1c} <7) and abnormal (HbA_{1c} ≥7). High fasting plasma glucose is considered, if the value is >130 mg/dl. Anemia is defined as a hemoglobin concentration of <13 g/dl in men and <12 g/dl in women and presence of nephropathy is considered microalbuminuria if albumin creatinine ratio (ACR) is between 30 and 300 mg/g and macroalbuminuria if ACR is above 300 mg/g, respectively.

The five methods described in Section 2 have been applied to this data set; based on the objective (1), we have attempted a method without rounding the digits intermittently and have considered this approach as a SM for comparison.

Out of the 1329 individuals with diabetes, 124 (9.33%) are DR. Table 2 demonstrates the results using logistic regression models with DR as dependent variable. The model fit is found to be satisfactory with overall

accuracy of 91.1% along with Hosmer and Lemeshow value ($\chi^2=2.956$, 8 d.f.) at 0.937 significance with the sensitivity 12.9% and Specificity 99.2%. Out of 12 factors, 11 are found to be risk factors for DR with statistically significant ($p<0.0001$) and BMI (OR: 0.37, $p<0.0001$) as protective factor of DR. The weight selection for M_2 , M_3 , M_4 and M_5 select the regression coefficient (0.304), mean of (0.304 and 0.308), highest Wald statistics corresponded regression coefficient (1.249) and highest regression coefficient (1.662) respectively. Table 3 shows the regression coefficients for the Table 2 and the points allocated to each risk factor category by difference methods for SN-DREAMS III data.

Table 4 shows that example of few patient profiles for different methods of risk estimation. SM yields exact risk estimation and other five methods (M_1 - M_5) provided subjective risk estimation. The patient 1 show percentage change from SM is uniformly higher in M_1 , M_2 and M_3 ; especially M_1 has highest up to 6.9%, M_2 has -12.6% and M_3 has -11.1% such differences highlights the effect of selection of weights, intermittent rounding the digits. Even the patient profile has the effect of high risk factors (FBS, Insulin, HbA_{1c} and History of family diabetes) M_2 and M_3 tend to decrease the risk when compared to the SM. In particular patients with all presence of all risk factors also yield a negative difference in M_2 . These observations are pictorially represented in Figure 1.

Table 5 presents the summary of difference existing between the methods (M_1 - M_5) compared with SM using the simulated combinations of 2¹² possibilities. And the percentage of deviation from the SM for all

Risk factors	β coefficient	S.E.	Wald	p	OR	OR 95% of CI	
						Lower	Upper
Intercept	-5.281	0.019	79293	<0.0001	0.005		
Age group (<55=0; ≥ 55=1)	0.308	0.008	1473	<0.0001	1.36	1.34	1.38
Gender (female=0; male=1)	0.619	0.008	6058	<0.0001	1.86	1.83	1.89
Duration of DM (<5= 0; ≥ 5=1)	1.249	0.008	23328	<0.0001	3.49	3.43	3.54
History of family diabetes (no=0; yes=1)	0.598	0.008	5542	<0.0001	1.82	1.79	1.85
User of insulin (no=0; yes=1)	1.662	0.018	8414	<0.0001	5.27	5.09	5.46
History of hypertension (no=0; yes=1)	0.861	0.008	11869	<0.0001	2.36	2.33	2.40
BMI (non-obese=0; obese=1)	-1.003	0.017	3326	<0.0001	0.37	0.35	0.38
Physical activity (heavy=0; moderate and less=1)	0.508	0.015	1131	<0.0001	1.66	1.61	1.71
FBS value (<130=0; ≥ 130=1)	0.658	0.010	3974	<0.0001	1.93	1.89	1.97
HbA _{1c} (≤7=0; >7=1)	0.677	0.009	5451	<0.0001	1.97	1.93	2.00
Anemia (no=0; yes=1)	0.304	0.010	937	<0.0001	1.36	1.33	1.38
Nephropathy (no=0; yes=1)	0.820	0.008	10884	<0.0001	2.27	2.24	2.31

DM- Diabetes Mellitus; BMI- Body Mass Index; FBS- Fasting Blood Sugar; HbA_{1c}- Glycated Hemoglobin

Table 2: Beta coefficients from the multiple logistic regression final model predicting diabetic retinopathy for SN-DREAMS III data.

Factors	Score allocated for difference methods					
	SM	M_1	M_2	M_3	M_4	M_5
Age Group (≥ 55 years)	0.308	3	2	1	25	19
Gender (male)	0.619	6	4	2	50	37
Duration of DM (≥ 5 years)	1.249	12	8	4	100	75
History of family diabetes (yes)	0.598	6	4	2	48	36
User of insulin (yes)	1.662	17	11	5	133	100
History of hypertension (yes)	0.861	9	6	3	69	52
BMI (obese)	-1.003	-10	-7	-3	-80	-60
Physical Activity (moderate and less)	0.508	5	3	2	41	31
FBS value (≥ 130)	0.658	7	4	2	53	40
HbA _{1c} (>7)	0.677	7	4	2	54	41
Anemia (yes)	0.304	3	2	1	24	18
Nephropathy (yes)	0.820	8	5	3	66	49

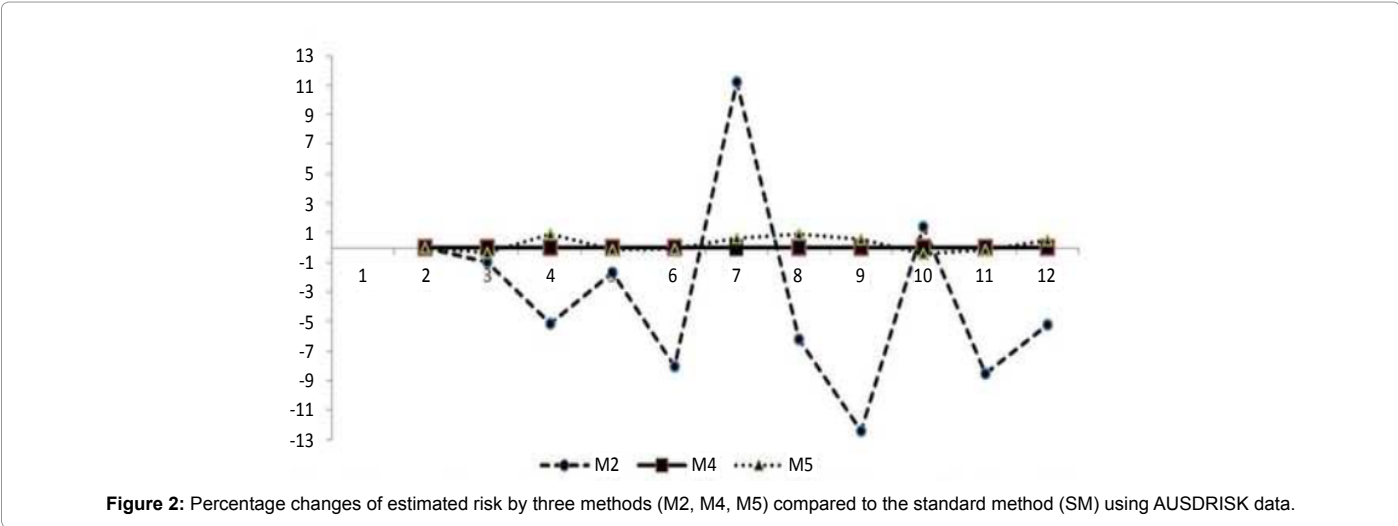
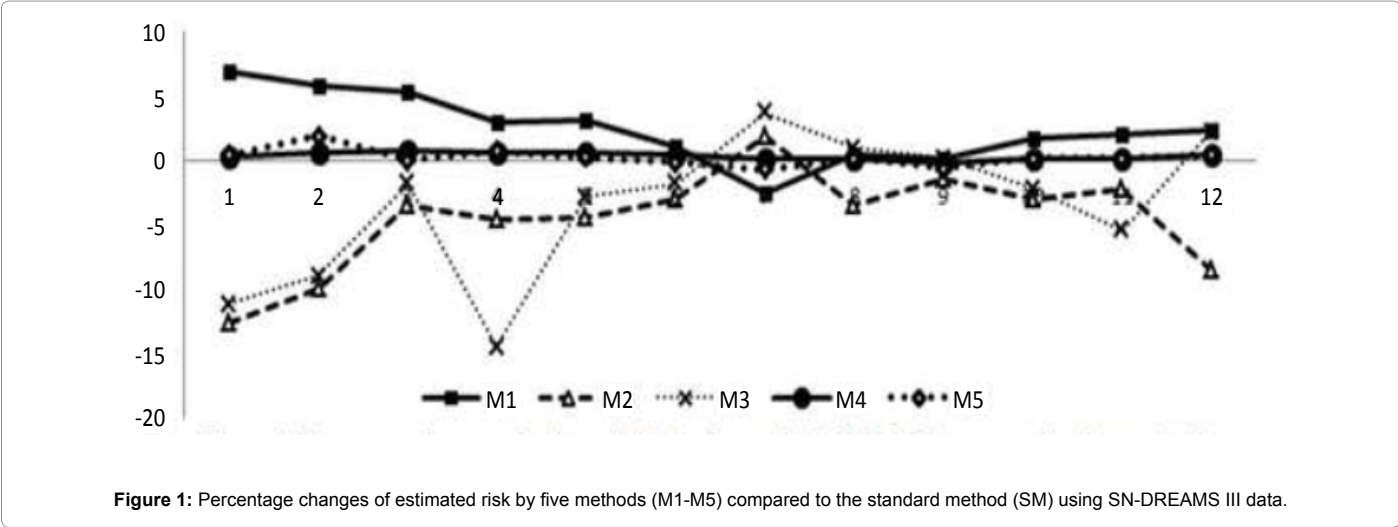
DM- Diabetes Mellitus; BMI- Body Mass Index; FBS- Fasting Blood Sugar; HbA_{1c}- Glycated Hemoglobin

Table 3: The points allocated to each component of the SN-DREAMS III score.

Example of few patient profile	Patients information												Risk estimation of difference method for each patients						Percentage change from SM				
	Age group (<55=0 ; ≥ 55=1)	Gender (female=0 ; male=1)	Duration of DM (< 5 = 0 ; ≥ 5=1)	History of family diabetes (no=0 ; yes=1)	User of insulin (no=0 ; yes=1)	History of hypertension (no=0 ; yes=1)	BMI (non obese = 0; obese=1)	Physical activity (heavy =0; moderate and less=1)	FBS value (< 130=0; ≥ 130 =1)	HbA1c (≤7-0=0; >7 =1)	Anemia (no=0 ; yes=1)	Nephropathy (no=0 ; yes=1)	SM	M ₁	M ₂	M ₃	M ₄	M ₅	M ₁	M ₂	M ₃	M ₄	M ₅
Patient 1	0	0	0	0	1	0	0	0	1	1	0	1	18.8	20.1	16.4	16.7	18.8	18.9	6.9	-12.6	-11.1	0.3	0.4
Patient 2	1	0	0	1	0	0	0	0	1	1	0	0	4.6	4.8	4.1	4.2	4.6	4.6	5.8	-9.9	-8.9	0.6	1.9
Patient 3	0	1	0	0	1	1	0	0	1	0	0	1	34.1	35.9	32.9	33.4	34.3	34.0	5.3	-3.5	-1.8	0.8	0.0
Patient 4	1	1	0	1	1	0	0	0	0	1	0	0	19.5	20.1	18.6	16.7	19.6	19.6	3.0	-4.6	-14.4	0.6	0.7
Patient 5	1	1	0	0	1	1	0	0	0	1	0	1	41.7	43.0	39.9	40.6	42.0	41.9	3.1	-4.4	-2.8	0.6	0.3
Patient 6	0	1	1	0	1	1	0	0	1	0	0	1	64.3	65.0	62.3	63.1	64.6	64.2	1.1	-3.0	-1.8	0.4	-0.1
Patient 7	0	1	1	1	0	1	0	0	0	0	1	0	16.1	15.7	16.4	16.7	16.1	16.0	-2.6	1.9	3.7	0.2	-0.7
Patient 8	1	1	1	1	1	1	1	1	1	1	1	1	87.9	88.3	84.8	88.8	88.1	88.1	0.5	-3.4	1.0	0.2	0.2
Patient 9	0	0	1	0	1	1	0	0	0	0	1	1	40.5	40.6	39.9	40.6	40.5	40.2	0.2	-1.4	0.2	-0.1	-0.6
Patient 10	0	0	1	1	1	1	0	0	1	1	0	1	77.6	78.9	75.3	75.9	77.7	77.7	1.7	-3.0	-2.2	0.1	0.2
Patient 11	0	1	1	1	1	1	0	0	1	1	0	0	73.9	75.4	72.3	69.9	74.1	74.1	2.0	-2.2	-5.4	0.2	0.2
Patient 12	0	0	1	1	0	1	0	0	1	1	0	1	39.7	40.6	36.3	40.6	39.9	39.8	2.3	-8.5	2.2	0.4	0.4

DM- Diabetes Mellitus; BMI- Body Mass Index; FBS- Fasting Blood Sugar; Hba1c- Glycated Hemoglobin

Table 4: The validating different methods of risk estimation and identifying percentage change from SM using SN-DREAMS III data.



Parameters	Percentage changes from SM in total 2 ¹² combinations				
	M ₁	M ₂	M ₃	M ₄	M ₅
N	4096	4096	4096	4096	4096
Minimum	-9.33	-25.04	-19.82	-7.8	-1.65
Maximum	14.22	7.10	41.50	2.46	3.50

Table 5: Minimum and maximum value for percentage change from SM in total 2¹² combinations.

Methods	Cronbach's Alpha	Intra class Correlation		Area under ROC	Bland-Altman test
		Estimate	95% of Confidence Interval		
M ₁	0.9997	0.9995	0.9994	0.8059	6.30%
M ₂	0.9963	0.9926	0.9917	0.8043	7.00%
M ₃	0.9986	0.9972	0.9969	0.8024	6.30%
M ₄	0.9999	0.9999	0.9999	0.8098	4.70%
M ₅	0.9999	0.9999	0.9999	0.8076	5.50%

Table 6: Results from validation analysis for five methods (M₁- M₅).

Factors	Score allocated for difference methods			
	SM	M ₂	M ₄	M ₅
Male sex	0.586	3	42	35
Age group(35-44)	0.455	2	32	27
Age group(45-54)	0.919	4	65	54
Age group(55-64)	1.3	6	92	77
Age group(≥65)	1.645	8	117	97
SE*	0.418	2	30	25
Parental history of diabetes	0.624	3	44	37
History of high blood glucose	1.358	6	96	80
Use of antihypertensive medications	0.462	2	33	27
Current smoker	0.463	2	33	27
Physical inactivity	0.428	2	30	25
Category 2 (WC)	0.884	4	63	52
Category 3 (WC)	1.411	7	100	83
Overweight (25-< 30)	0.569	3	40	34
Obese (30-< 35)	1.224	6	87	72
Morbidly obese (≥35)	1.698	8	120	100

Table 7: The points allocated to each component of the AUSDRISK score.

other the five methods show that maximum range is goes up to 41.50% in M₃ compared to M₄ (2.46%) and M₅ (3.50%). Similarly, the lower range for M₂ is -25.04 compared to M₄ (-0.78%) and M₅ (-1.65%). Table 6 specifies the agreement of M₁-M₅ with the SM where M₄ has the highest Cronbach's alpha and intra class correlation followed by that of M₅. Further depicts the nature of outliers identified using Bland-Altman test from which it can be shown that M₂ has maximum outliers (7.0%), followed by M₁ and M₃ (6.3%), M₅ (5.5%) whereas least is M₄ (4.74%). This observation is further augmented by the ROC using a real data set, DR as classifier with the risk estimation of M₁-M₅ which illuminates that M₄ and M₅ are the top 2 values compared to other three methods.

Now, the M₄ and M₅ are found to be better than the existing methodology. In order to validate, those procedures are applied to another set of information from the research article AUSDRISK [14] to understand the effect of M₄ and M₅ compared to M₂ originally adopted. Table 7 shows the regression coefficients from the article AUSDRISK and the points allocated to each risk factor category for the M₄ and M₅ are computed. Then, the Table 8 presents the percentage of changes from the SM using combinations and Figure 2 depicts the pictorial form of percentage change compared to standard method. It can observe from Table 5 that M₄ and M₅ have uniformly lesser difference composed to that of M₂. In particular this cross validation has shown that M₄ has a difference in an order of 10⁻⁶ to 10⁻¹⁰. However the difference to M₂ has gone up to 11.2% in positive side and -8.5% in the negative side.

Conclusion

The present study has considered three (M₁, M₂ and M₃) popular methods for estimating risk for the patients of a particular disease. All these methods utilize extensively the regression coefficients of logistic regression model. Since setting score cards is considered as population disease specific, a need has been felt to establish suitable procedure, understand the limitation of existing procedures through a planned study design.

Example of few patient profile	Patients information															Risk estimation of difference method for each patients				Percentage change from SM			
	Male sex	Age group(35-44)	Age group(45-54)	Age group(55-64)	Age group(≥65)	SE*	Parental history of diabetes	History of high blood glucose	Use of antihypertensive medications	Current smoker	Physical inactivity	Category 2 (WC)	Category 3 (WC)	Overweight (25-< 30)	Obese (30-< 35)	Morbidly obese (≥35)	SM	M ₂	M ₄	M ₅	M ₂	M ₄	M ₅
Patient 1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.46	0.46	0.46	0.46	0.00	0.00	0.00
Patient 2	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0.70	0.69	0.70	0.70	-0.99	<0.0001	-0.35
Patient 3	1	1	0	0	0	0	1	0	0	1	0	0	0	0	0	1	17.39	16.49	17.39	17.56	-5.18	<0.0001	0.95
Patient 4	1	0	0	0	0	0	0	0	1	1	0	0	0	0	1	0	6.61	6.49	6.61	6.60	-1.67	<0.0001	-0.11
Patient 5	0	0	0	0	1	0	0	1	1	0	0	0	0	0	0	1	44.50	40.90	44.50	44.47	-8.08	<0.0001	-0.06
Patient 6	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	3.22	3.58	3.22	3.24	11.21	<0.0001	0.65
Patient 7	1	0	0	1	0	1	0	1	1	0	0	0	0	0	1	0	49.10	46.03	49.10	49.54	-6.25	<0.0001	0.90
Patient 8	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	1	26.35	23.08	26.35	26.50	-12.41	<0.0001	0.58
Patient 9	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	1	13.62	13.81	13.62	13.56	1.39	<0.0001	-0.44
Patient 10	0	0	0	0	0	0	1	1	1	1	1	1	0	0	1	0	61.70	56.44	61.70	61.64	-8.53	<0.0001	-0.11
Patient 11	1	0	0	1	0	1	0	1	1	1	1	1	0	1	0	0	74.73	70.81	74.73	75.07	-5.24	<0.0001	0.46
Patient 12	0	0	0	0	0	1	1	0	1	0	1	1	1	0	0	0	23.92	23.08	23.92	23.94	-3.53	<0.0001	0.08

*-SE (Southern European, Asian, Aboriginal and Torres Strait Islander or Pacific Islander background)

Table 8: The validating different methods of risk estimation and identifying percentage change from SM using AUSDRISK data.

Existing methods are not effective to capture the risk estimation due to various barriers. This study has aimed to consolidate few widely applied methods on risk score systems and device two new methods (M_4 and M_5). Further, standard method (SM) has been included in the study that is based on non-rounding the digits in the interim computations. This is extended to a comparative study with the existence and two different methods (M_4 and M_5).

In M_4 coefficients that indicate the higher significance risk factors among statistically significant variables using Wald Statistics are obtained from the output of logistic regression modeling. Also in the method M_5 the highest positive regression coefficient has been selected for score calculation. Both notions are to strengthen the risk score calculations that are indicated by the underlying variables of the regression model. All these procedures are applied to a cross-sectional study to estimate the prevalence and risk factors for diabetic retinopathy in rural south India, SN-DREAMS III.

From the results it could be observed that methods M_4 and M_5 are more appropriate in the score card development for identifying the risk factors involved in DR. The result have further indicated that

- Notable differences exist among the methods M_1 , M_2 , M_3 , but not in M_4 , M_5 .
- M_3 records the highest difference and M_4 has the least.
- Such differences yield a reversal estimation patterns within M_2 or M_3 .
- High risk patient's profile has been estimated as a lower level and vice versa, when M_2 and M_3 are applied.

Also the methods have been cross validated using another data set from AUSDRISK [14] that also supports the above observations. By considering few statistical properties such as the choice weights, sampling techniques, and validation procedures, the present work has observed that M_4 and M_5 could be more suitable methods for score estimation in diseases like diabetic retinopathy. Similar attempts would help to investigate the usefulness of methods for risk score estimation involved in other diseases and different geographic locations.

References

1. Golubnitschaja O (2010) Advanced diabetes care: three levels of prediction, prevention personalized treatment. *Curr Diabetes Rev* 6: 42-51.
2. Anjana RM, Pradeepa R, Deepa M, Datta M, Sudha V, et al. (2011) Prevalence of diabetes and prediabetes (impaired fasting glucose and/or impaired glucose tolerance) in urban and rural India: phase I results of the Indian Council of Medical Research-India Diabetes (ICMR-INDIAB) study. *Diabetologia* 54: 3022-3027.
3. Ramachandran A, Snehalatha C, Ma RC (2014) Diabetes in South-East Asia: An update. *Diabetes Res Clin Pract* 103: 231-237.
4. Nam HC, David W, Leonor G, Pablo AM, Nita F, et al. (2013) IDF Diabetes Atlas. *International Diabetes Federation* 6: 34.
5. Cheung N, Mitchell P, Wong TY (2010) Diabetic retinopathy. *The Lancet* 376: 124-136.
6. Zhang X, Saaddine JB, Chou C-F, Cotch MF, Cheng YJ, et al. (2010) Prevalence of diabetic retinopathy in the United States, 2005-2008. *JAMA* 304: 649-656.
7. Askew DA, Crossland L, Ware RS, Begg S, Cranston P, et al. (2012) Diabetic retinopathy screening and monitoring of early stage disease in general practice: design and methods. *Contemp Clin Trials* 33: 969-975.
8. Hosseini SM, Maracy MR, Amini M, Baradaran HR (2009) A risk score development for diabetic retinopathy screening in Isfahan-Iran. *J Res Med Sci* 14: 105-110.
9. Han Y, Schneck ME, Bearse MA Jr, Barez S, Jacobsen CH, et al. (2004) Formulation and evaluation of a predictive model to identify the sites of future diabetic retinopathy. *Invest Ophthalmol Vis Sci* 45: 4106-4112.
10. McNab H, Wynn A (2004) Principles and practice of consumer credit risk management. CIB Publishing, Canterbury.
11. Glümer C, Carstensen B, Sandbaek A, Lauritzen T, Jørgensen T, et al. (2004) A Danish diabetes risk score for targeted screening: the Inter99 study. *Diabetes Care* 27: 727-733.
12. Glümer C, Vistisen D, Borch-Johnsen K, Colagiuri S, DETECT-2 Collaboration (2006) Risk scores for type 2 diabetes can be applied in some populations but not all. *Diabetes Care* 29: 410-414.
13. Wu C, Hannan EL, Walford G, Ambrose JA, Holmes DR Jr, et al. (2006) A risk score to predict in-hospital mortality for percutaneous coronary interventions. *J Am Coll Cardiol* 47: 654-660.
14. Chen L, Magliano DJ, Balkau B, Colagiuri S, Zimmet PZ, et al. (2010) AUSDRISK: an Australian Type 2 Diabetes Risk Assessment Tool based on demographic, lifestyle and simple anthropometric measures. *Med J Aust* 192: 197-202.
15. Sugioka T, Hayashino Y, Konno S, Kikuchi S, Fukuhara S (2008) Predictive value of self-reported patient information for the identification of lumbar spinal stenosis. *Fam Pract* 25: 237-244.
16. Sullivan LM, Massaro JM, D'Agostino RB (2004) Presentation of multivariate data for clinical use: the Framingham Study risk score functions. *Stat Med* 23: 1631-1660.
17. Hosmer DW, Hjort NL (2002) Goodness-of-fit processes for logistic regression: simulation results. *Stat Med* 21: 2723-2738.
18. Hanley JA, McNeil BJ (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143: 29-36.
19. Bland JM, Altman DG (1995) Comparing methods of measurement: why plotting difference against standard method is misleading. *The Lancet* 346: 1085-1087.
20. Bland JM, Altman DG (1997) Statistics notes: Cronbach's alpha. *BMJ* 314: 572.
21. McGraw KO, Wong SP (1996) Forming inferences about some intraclass correlation coefficients. *Psychological Methods* 1: 30-46.
22. Pal SS, Raman R, Ganesan S, Sahu C, Sharma T (2011) Sankara Nethralaya Diabetic Retinopathy Epidemiology and Molecular Genetic Study (SN-DREAMS III): Study design and research methodology. *BMC Ophthalmol* 11: 1-7.