ORIGINAL ARTICLE

IDENTIFYING NEWLY DIAGNOSED DIABETES MELLITUS RISK FACTORS USING GRAPHICAL NETWORKS WITH EXPERT KNOWLEDGE

Ooi Wei Lim[1*] and Sarat C Dass[2]

[1]Heriot-Watt Global College Malaysia, Heriot-Watt University, Malaysia
[2]School of Mathematical and Computer Science, Heriot-Watt University, Malaysia

***Corresponding author: Lim Ooi Wei**
E-mail address: o.lim@hw.ac.uk

ABSTRACT

An increasing trend of newly diagnosed diabetes mellitus (DM) among adolescents is occurring worldwide, including Malaysia. This study aims to determine the overall relationships between risk factors on the prevalence of newly diagnosed diabetes mellitus among Malaysian adolescents. Current study uses a cross-sectional study, data from the Fifth National Health and Morbidity Survey 2015 which consists of individuals who ages 18 and above, extraction of 18 and 19 years old from the data set was done.. Bayesian networks modelling was performed by using graphical networks with expert knowledge to identify the risk factors of newly diagnosed diabetes mellitus among adolescents in Malaysia. Education levels, Body Mass Index (BMI), and physical inactivity were identified as the significant predictors of newly diagnosed DM. The highest conditional probability of developing newly diagnosed DM belongs to both obese and underweight respondents given they have no formal education (probability, pr = 0.5000), followed by obese respondents who had unclassified level of education (pr = 0.4709), and obese respondents with primary level of education (pr = 0.4692). The findings of current study provide insights and allows policymakers to plan for future interventions in order to monitor and reduce the high prevalence of newly diagnosed diabetes mellitus among adolescents in Malaysia.

***Keywords**: adolescents, newly diagnosed diabetes mellitus, behavioral risk factors, sociodemographic risk factors*

INTRODUCTION

Diabetes mellitus (DM) is a metabolic disease caused by increased blood glucose or blood sugar. If left untreated over time, it may cause damage to several vital body parts such as the heart, blood vessels, eyes, kidneys and nerves that may increase the risk of premature death in patients with diabetes[1-3]. Despite the interventions Malaysian government, the prevalence of newly diagnosed diabetes mellitus increased from 8.0% (95% CI: 7.2 - 8.7) to 9.2% (95% CI: 8.5, 9.9) from the year 2011 to 2015. A similar trend was found in the age group 18-19 years, of which the prevalence increased from 2.1% (95% CI: 1.2 - 3.7) to 5.5% (95% CI: 3.9, 7.7) for the year 2011 and 2015 respectively [4,5]. Therefore, it is essential to address this issue among adolescents because prevention is better than cure.

The risk factors for diabetes mellitus (DM) can be categorized into two groups: behavioral and sociodemographic risk factors. According to NHMS 2015 report, the overall prevalence of physically active adults was 66.5% (95% CI: 65.6, 67.6). Males [71.1% (95% CI: 69.7, 72.4] were found significantly more active than females [61.7 (95% CI: 60.3, 63.2].Previous studies showed that the

lack of physical activity is prevalent and have higher percentage than people who tends to exercise[6-8]. In addition, people with obesity (BMI >= 30.00 kg/m2) are more likely to have newly diagnosed diabetes mellitus [7,9-11]. In 2004, a study reported the prevalence of overweight and obesity among Malaysian school children and adolescents to be 7.3% [12].

On the other hand, sociodemographic risk factors such as household location was shown to be independently associated with having newly diagnosed diabetes mellitus whereby persons living outside of major city were more likely to have newly diagnosed diabetes mellitus [13]. This may reflect poorer access to health services for those in lower socioeconomic groups and those who reside in areas outside of major city. Besides, education has an inverse correlation with the prevalence of newly diagnosed diabetes mellitus where previous studies suggested that people with lower education levels are more susceptible to newly diagnosed diabetes mellitus ($P < 0.01$) [14-16]. Multiple studies suggested that there was a significant difference (overall $P < 0.001$) in prevalence of newly diagnosed diabetes mellitus among the ethnic group, of which non-Hispanic Asian groups (all non-Hispanic persons who did not self-identify as black and had origins in any of the Asian countries) were the highest among other ethnicity groups [14,15].

Majority of the studies available were conducted on a one-to-one relationship basis between risk factors and the prevalence of non-communicable

diseases, but no discussion on the possibility of conditional distributions and the interrelationships between the behavioral and sociodemographic risk factors on the prevalence of non-communicable diseases. For instance, obesity is one of the behavioral risk factors of diabetes mellitus, yet it can be a consequence of physical inactivity, which eventually leads to higher prevalence of newly diagnosed diabetes mellitus. Besides, there has been no study regarding to the risk factors of newly diagnosed diabetes mellitus among adolescents in Malaysia.

Therefore, this study aimed to investigate the overall relationships of risk factors of newly diagnosed diabetes mellitus among adolescents in Malaysia. The findings of this study would enhance the understanding and awareness between the conditional probabilities and the interrelationships among the sociodemographic and behavioral risk factors on the prevalence of newly diagnosed diabetes mellitus among Malaysian adolescents which ultimately would assist the Government to strategize appropriate intervention in reducing and monitoring the prevalence of newly diagnosed diabetes mellitus especially among adolescents in Malaysia.

METHODS

Study subjects and design

Current study is a secondary data analysis of the National Health and Morbidity Survey 2015 (NHMS 2015). The secondary data of this study was obtained through application from the Ministry of Health (MOH). The NHMS 2015 is a cross-sectional study that used two-stage stratified random sampling to ensure sample collected would receive nationwide responses and represent population. Targeted population for this survey includes people of age 18 and above. The first stratum includes the states of Malaysia in addition of the federal territories. Furthermore, the second stratum involves the urban and rural strata that formed within the first stratum. Institutional population such as hotel, hostels, hospital, etc. were excluded from the survey. NHMS 2015 applied an updated sampling frame provided by Department of Statistics Malaysia. 10,428 Living Quarters were selected from 536 urban Enumeration Blocks (EBs) and 333 rural EBs [17].

Face-to-face interview and self-administered structured questionnaires were employed to carry out data collection of the survey. With consent of respondents, trained nurses will carry out clinical examinations and measure the blood pressure, height, weight, waist and calf circumferences, and a blood sampling via finger prick test to check level of cholesterol, glucose and hemoglobin. Current study is registered under the National Medical Research Registry (NMRR-19-3259-50433).

Variables

Dependent variable

Current study explores the prevalence and risk factors of newly diagnosed diabetes mellitus among Malaysian adolescents. Based on NHMS 2015 report, newly diagnosed diabetes mellitus is defined as not known to have diabetes and had a fasting capillary blood glucose (FBG) of 6.1 mmol/L or more (or non-fasting blood glucose of more than 11.1 mmol/L)[5]. Questionnaire and CardioChek portable blood test system were used to measure the finger-pricked fasting blood glucose in the diabetes survey.

A total of 19,935 respondents responded to the diabetes questionnaire meanwhile only 16,361 out of 17,809 respondents who claimed that they were not diabetic consented for the finger-prick test[5]. Per definition set by World Health Organization (WHO), adolescents are individuals of age 10-19 years old. Refined sample data contains 803 adolescents but 765 respondents have consented and completed the survey in full.

Independent variables

All independent variables are categorical data and classified into two (2) main groups, behavioral risk factors and sociodemographic risk factors. The behavioral risk factors include:

(i) Body Mass Index (BMI) and it is classified using World Health Organization (1998) guideline with four categories: underweight ($< 18.5\ kg/m^2$), normal ($18.5 - 24.99\ kg/m^2$), Overweight ($25.0 - 29.99\ kg/m^2$) and obese ($\geq 30 kg/m^2$).

(ii) Smoking status of respondents were evaluated by asking about their smoking habit and favorite tobacco product if applicable. Results obtained were categorized into current tobacco smoker and current non-smoker. Based on the NHMS 2015 report, current smoker was defined as smoker who daily or occasionally smokes any tobacco product .

(iii) Physical activity of respondents were tested with short version of International Physical Activity Questionnaire (IPAQ). If the respondents achieve at least a minimum of 600 MET-min/week, they will be classified as active else they will be classified as inactive.

On the other hand, the sociodemographic risk factors included in current study are gender, strata (rural or urban), ethnicity (Malay; Chinese; Indian; Others Bumiputra; Others), education levels (no formal education; primary; secondary; tertiary; unclassified), state of origin of respondents.

Statistical analysis

Prior to statistical modelling, data cleaning was carried out to remove any anomalies found to provide complete and error free information. Clean data can provide better decision making process and give better prediction which lead to an increased productivity [18,19]. Descriptive statistics of dataset were analyzed to determine the frequency and distribution of newly diagnosed diabetes mellitus among Malaysian adolescents. Pearson's chi-square test of independence is used to determine whether there is an association between the variables with respect to newly diagnosed diabetes mellitus. A conditional independence test is used to evaluate if probabilistic dependency has been supported by the data. Variables that passed the test are used to construct the arc of Bayesian network.

Pearson's χ^2 test has been conducted to assess the significance of association of newly diagnosed diabetes mellitus and different risk factors in order to select important risk factors. A combination of data-driven approach and experts' knowledge were used to construct the Bayesian network models. Confidence independence (CI) test was carried out for risk factors selection to infer the edges between nodes in the Bayesian Network graph. Based on the CI tests, expert knowledge has been employed to come up with candidate choices of the Bayesian Network (BN) structure. Hence, data-driven approach is used to construct the basic model and experts' knowledge is used to further improve the model design. Goodness-of-fit of each model is tested via network scores by measuring how well the directed acyclic graph of Bayesian network mirrors the dependence structure of data[20]. Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC) are used to calculate the network score, model with the lowest scores in both criterion will be chosen for current study[20]. Bayesian networks are a type of probabilistic graphical model that uses Bayesian inference for probability computation. The level of significance is defaulted to 0.05 throughout the study. All statistical analysis were conducted using R version 4.0.2 (R Core Team, 2020) and the $bnlearn\ (v4.5; Marco\ Scutari, 2019)$ package.

RESULTS

Descriptive statistics

NHMS 2015 has collected 19,935 responses over the survey period. However, upon filtering the data based on adolescent age group, we have extracted 765 samples of adolescent respondents of age 18 to 19 years old, who completed the survey and clinical examinations. Table 1 shows majority of respondents (94%) were identified as non-diabetic, whereas a small portion of the respondents (6%) were identified to have newly diagnosed diabetes mellitus.

Table 1: Frequency of Diabetes Mellitus

Diabetes Mellitus (DM)	Frequency	Percentage (%)
No DM	719	94.00
Undiagnosed DM	46	6.00
Total	765	100.00

A two-way relationship analysis between risk factors of respondents and their respective diabetes mellitus status was carried out and results are shown in Table 2. Male adolescents showed higher percentage (3.4%) of newly diagnosed diabetes mellitus in comparison to the female adolescents (2.6%). With respect to education, respondents with secondary education level show higher relative frequency of having newly diagnosed diabetes mellitus (5.6%) compared to other educational level. This is followed by respondents with no formal education (0.1%), primary education (0.1%), as well as tertiary education (0.1%). The survey reported that there was no cases of newly diagnosed diabetes mellitus among respondents with unclassified education level. With respect to ethnicity, Malay respondents show highest relative frequency (3.8%) of having newly diagnosed diabetes mellitus compare to others. Other Bumiputra show second highest relative frequency (0.9%) and is closely followed by Indian (0.8%) whereas Chinese and Other ethnicity have the lowest relative frequency (0.3%) of having newly diagnosed diabetes mellitus.

With regards to strata , respondents from both rural and urban strata exhibited same relative frequency (3.0%) of having newly diagnosed diabetes mellitus. Next, with regard to respondent's state of origin, Johor exhibits the highest relative frequency (1.7%) of having newly diagnosed diabetes mellitus as compared to other states. On the other hand, Kedah, Labuan, Melaka, WP Kuala Lumpur and WP Putrajaya have zero cases of newly diagnosed diabetes mellitus among adolescents.

In terms of body mass index (BMI), respondents with normal BMI showed the highest relative frequency of having newly diagnosed diabetes mellitus (3.9%), followed by overweight (0.8%) and obese (0.8%). The lowest relative frequency

of having newly diagnosed diabetes mellitus in terms of BMI is associated with underweight adolescents (0.5%). Current study shows that current non-smokers display a higher relative frequency (4.0%) of having newly diagnosed diabetes mellitus as compared to current tobacco smokers (2.0%). With respect to physical activity, adolescents who are physically active show higher relative frequency (4.6%) of having newly diagnosed diabetes mellitus as compared to physically inactive adolescents (1.4%).

Table 2: Summary of respondents' information

Variables	Level(s)	No DM Frequency (%)	Newly diagnosed DM Frequency (%)
Gender	Male	371 (48.5)	26 (3.4)
	Female	348 (45.5)	20 (2.6)
Education level	Unclassified	23 (3.0)	0 (0)
	No formal education	8 (1.0)	1 (0.1)
	Primary education	36 (4.7)	1 (0.1)
	Secondary education	570 (74.5)	43 (5.6)
	Tertiary education	82 (10.7)	1 (0.1)
Body Mass Index (BMI)	Underweight	147 (19.2)	4 (0.5)
	Normal	376 (49.2)	30 (3.9)
	Overweight	104 (13.6)	6 (0.8)
	Obese	92 (12.0)	6 (0.8)
Ethnicity	Malay	469 (61.3)	29 (3.8)
	Chinese	100 (13.1)	2 (0.3)
	Indian	50 (6.5)	6 (0.8)
	Others	34 (4.4)	2 (0.3)
	Other Bumiputra	66 (8.6)	7 (0.9)
Smoking status	Current tobacco smoker	142 (18.6)	15 (2.0)
	Current non-smoker	577 (75.4)	31 (4.0)
Physical activity	Active	448 (58.6)	35 (4.6)
	Inactive	271 (35.4)	11 (1.4)
Strata	Urban	440 (57.5)	23 (3.0)
	Rural	279 (36.5)	23 (3.0)
State of Origin (NG_new)	Johor	58 (7.58)	13 (1.70)
	Kedah	51 (6.70)	0 (0.00)
	Kelantan	56 (7.30)	3 (0.39)
	Labuan	1 (0.10)	0 (0.0)
	Melaka	30 (3.90)	0 (0.0)
	Negeri Sembilan	37 (4.80)	2 (0.26)
	Pahang	43 (5.60)	1 (0.13)
	Penang	43 (5.60)	2 (0.26)
	Perak	52 (6.80)	3 (0.39)
	Perlis	32 (4.20)	5 (0.65)
	Sabah	56 (7.30)	6 (0.78)
	Sarawak	50 (6.50)	1 (0.13)
	Selangor	137 (17.90)	6 (0.78)
	Terengganu	49 (6.40)	4 (0.52)
	WP Kuala Lumpur	20 (2.60)	0 (0.00)
	WP Putrajaya	4 (0.50)	0 (0.00)

Graphical reasoning

Conditional independence tests were carried out to evaluate if probabilistic dependency has been supported by the data, result is shown in Table 3. With respect to the newly diagnosed diabetes mellitus risk factors, any combination of independent variables and conditional variables that shows p-value of 0.05 and above will be considered as part of the Bayesian network. Out of the 46 tested combinations, only 8 combinations have p-value less than 0.05, hence the exclusion in further modelling.

Table 3: Results of conditional independence test

Dependent variable	Independent variable	Conditional variable(s)	p-value
Newly diagnosed DM	BMI	Education	0.4318
		Ethnicity	0.5816
		NG_new	0.4318
		Physical activity	0.4921
		Smoking status	0.6365
		Strata	0.4979
Newly diagnosed DM	Ethnicity	BMI	0.4432
		Education	0.2007
		NG_new	0.2807
		Physical activity	0.3984
		Smoking status	0.0529
		Strata	0.3154
Newly diagnosed DM	Education	BMI	0.4733
		Ethnicity	0.2655
		NG_new	0.5805
		Physical activity	0.1798
		Smoking status	0.0464
		Strata	0.1459
Newly diagnosed DM	NG new	BMI	0.0989
		Education	0.0179
		Ethnicity	0.0411
		Physical activity	0.0464
		Smoking status	0.0885
		Strata	0.0885
Newly diagnosed DM	Physical activity	BMI	0.4181
		Education	0.2996
		Ethnicity	0.3414
		NG_new	0.3727
		Smoking status	0.0051
		Strata	0.1320
Newly diagnosed DM	Smoking status	BMI	0.5233
		Education	0.0489
		Ethnicity	0.0318
		NG_new	0.6127
		Physical activity	0.0057
		Strata	0.2231
Newly diagnosed DM	Strata	BMI	0.7773
		Education	0.3582
		Ethnicity	0.4214
		NG_new	0.7453
		Physical activity	0.2152
		Smoking status	0.3916

Initial model was created via data-driven approach and further improved by expert-driven approach. Figure 1 shows the finalized network geometry of the outcomes in this study that summarizes the prediction of sociodemographic and behavioral risk factors of newly diagnosed diabetes mellitus among Malaysian adolescents.

The goodness-of-fit statistics of the network are as follows:

$$AIC = -6654.519;\ BIC = -8969.519$$

The correlation between risk factors and newly diagnosed diabetes mellitus were quantified by fitted conditional probability calculated via Bayes' theorem. Therefore, conditional probability will edge closer to true value and differ from initial relative frequency shown in Table 2. Figure 1 shows that education, body mass index, states of origin (NG_new) and smoking status of respondents were significantly related to newly diagnosed diabetes mellitus, as such they can be good predictors.

Results of the conditional independence test among the risk factors and newly diagnosed diabetes mellitus have been tabulated. Based on the Bayesian network analysis, the highest conditional probability of having newly diagnosed diabetes mellitus is 0.500, which was found on respondents who have no formal education, given they are obese or underweight. It is the followed by respondents who have unclassified education

(0.4709) and primary education (0.4692), given they are obese. Based on Table 4, it is concluded that respondents with no formal education, unclassified education, and primary education face high probability of having newly diagnosed diabetes mellitus, given they suffer from overweight or obesity. Furthermore, conditioning on no formal education, respondents who are physically inactive have high conditional probability of having newly diagnosed diabetes mellitus (0.4493). With respect to smoking status of respondents, the conditional probability of having newly diagnosed diabetes mellitus differs upon condition on the state of origin

Table 4 a: Results of conditional probability test

Newly diagnosed DM	Node 1	Node 2	State 1	State 2	Conditional Probability
Yes	Education	BMI	No formal education	Underweight	0.5000
			Primary	Underweight	0.3590
			Secondary	Underweight	0.0499
			Tertiary	Underweight	0.2599
			Unclassified	Underweight	0.3685
Yes	Education	BMI	No formal education	Normal	0.3489
			Primary	Normal	0.2657
			Secondary	Normal	0.0807
			Tertiary	Normal	0.1080
			Unclassified	Normal	0.2934
Yes	Education	BMI	No formal education	Overweight	0.4478
			Primary	Overweight	0.3879
			Secondary	Overweight	0.0925
			Tertiary	Overweight	0.1868
			Unclassified	Overweight	0.4581
Yes	Education	BMI	No formal education	Obese	0.5000
			Primary	Obese	0.4692
			Secondary	Obese	0.0967
			Tertiary	Obese	0.1431
			Unclassified	Obese	0.4709
Yes	Ethinc3	BMI	Chinese	Underweight	0.0893
			Indian	Underweight	0.0893
			Malays	Underweight	0.0893
			Others	Underweight	0.0893
			Others Bumiputra	Underweight	0.0893
Yes	Ethinc3	BMI	Chinese	Normal	0.1035
			Indian	Normal	0.1035
			Malays	Normal	0.1035
			Others	Normal	0.1035
			Others Bumiputra	Normal	0.1035
Yes	Ethinc3	BMI	Chinese	Overweight	0.1314
			Indian	Overweight	0.1314
			Malays	Overweight	0.1314
			Others	Overweight	0.1314
			Others Bumiputra	Overweight	0.1314
Yes	Ethinc3	BMI	Chinese	Obese	0.1453
			Indian	Obese	0.1453
			Malays	Obese	0.1453
			Others	Obese	0.1453
			Others Bumiputra	Obese	0.1453
Yes	Ethinc3	Smoking status	Chinese	Current non-smoker	0.0884
			Indian	Current non-smoker	0.0884

Table 1b : Results of conditional probability test

Newly diagnosed DM	Node 1	Node 2	State 1	State 2	Conditional Probability
Yes	Ethinc3	Smoking status	Chinese	Current tobacco smoker	0.1988
			Indian	Current tobacco smoker	0.1988
			Malays	Current tobacco smoker	0.1988
			Others	Current tobacco smoker	0.1988
			Others Bumiputra	Current tobacco smoker	0.1988
Yes	Physical activity	BMI	Active	Underweight	0.0905
			Inactive	Underweight	0.0878
			Active	Normal	0.1109
			Inactive	Normal	0.0904
Yes	Physical activity	BMI	Active	Overweight	0.1199
			Inactive	Overweight	0.1661
Yes	Physical activity	BMI	Active	Obese	0.1299
			Inactive	Obese	0.1658
Yes	Physical activity	Education	Active	No formal education	0.3836
			Inactive	No formal education	0.4493
Yes	Physical activity	Education	Active	Primary	0.2996
			Inactive	Primary	0.3762
Yes	Physical activity	Education	Active	Secondary	0.0782
			Inactive	Secondary	0.0780
Yes	Physical activity	Education	Active	Tertiary	0.1559
			Inactive	Tertiary	0.1467
Yes	Physical activity	Education	Active	Unclassified	0.3520
			Inactive	Unclassified	0.3836
Yes	Smoking status	BMI	Current non-smoker	Underweight	0.0678
			Current tobacco smoker	Underweight	0.1789
Yes	Smoking status	BMI	Current non-smoker	Normal	0.0925
			Current tobacco smoker	Normal	0.1463
Yes	Smoking status	BMI	Current non-smoker	Overweight	0.0877
			Current tobacco smoker	Overweight	0.2989
Yes	Smoking status	BMI	Current non-smoker	Obese	0.1022
			Current tobacco smoker	Obese	0.3220
Yes	Smoking status	NG_new	Current non-smoker	Johor	0.1921
			Current tobacco smoker	Johor	0.2754

Table 4c: Results of conditional probability test

Newly diagnosed DM	Node 1	Node 2	State 1	State 2	Conditional Probability
Yes	Smoking status	NG_new	Current non-smoker	Kedah	0.0362
			Current tobacco smoker	Kedah	0.1474
Yes	Smoking status	NG_new	Current non-smoker	Kelantan	0.0918
			Current tobacco smoker	Kelantan	0.2208
Yes	Smoking status	NG_new	Current non-smoker	Labuan	NaN
			Current tobacco smoker	Labuan	0.2751
Yes	Smoking status	NG_new	Current non-smoker	Melaka	0.0528
			Current tobacco smoker	Melaka	0.2751
Yes	Smoking status	NG_new	Current non-smoker	Negeri Sembilan	0.1378
			Current tobacco smoker	Negeri Sembilan	0.1484
Yes	Smoking status	NG_new	Current non-smoker	Pahang	0.0249
			Current tobacco smoker	Pahang	0.1162
Yes	Smoking status	NG_new	Current non-smoker	Penang	0.0798
			Current tobacco smoker	Penang	0.1493
Yes	Smoking status	NG_new	Current non-smoker	Perak	0.0744
			Current tobacco smoker	Perak	0.1803
Yes	Smoking status	NG_new	Current non-smoker	Perlis	0.1516
			Current tobacco smoker	Perlis	0.3803
Yes	Smoking status	NG_new	Current non-smoker	Sabah	0.1095
			Current tobacco smoker	Sabah	0.3460
Yes	Smoking status	NG_new	Current non-smoker	Sarawak	0.0774
			Current tobacco smoker	Sarawak	0.1697
Yes	Smoking status	NG_new	Current non-smoker	Selangor	0.0525
			Current tobacco smoker	Selangor	0.1044
Yes	Smoking status	NG_new	Current non-smoker	Terengganu	0.0928
			Current tobacco smoker	Terengganu	0.4214
Yes	Smoking status	NG_new	Current non-smoker	WP Kuala Lumpur	0.0398
			Current tobacco smoker	WP Kuala Lumpur	0.2679
Yes	Smoking status	NG_new	Current non-smoker	WP Putrajaya	0.3907
			Current tobacco smoker	WP Putrajaya	0.2751
Yes	Strata	Smoking status	Rural	Current non-smoker	0.0884
			Urban	Current non-smoker	0.0884
Yes	Strata	Smoking status	Rural	Current tobacco smoker	0.1981
			Urban	Current tobacco smoker	0.1993

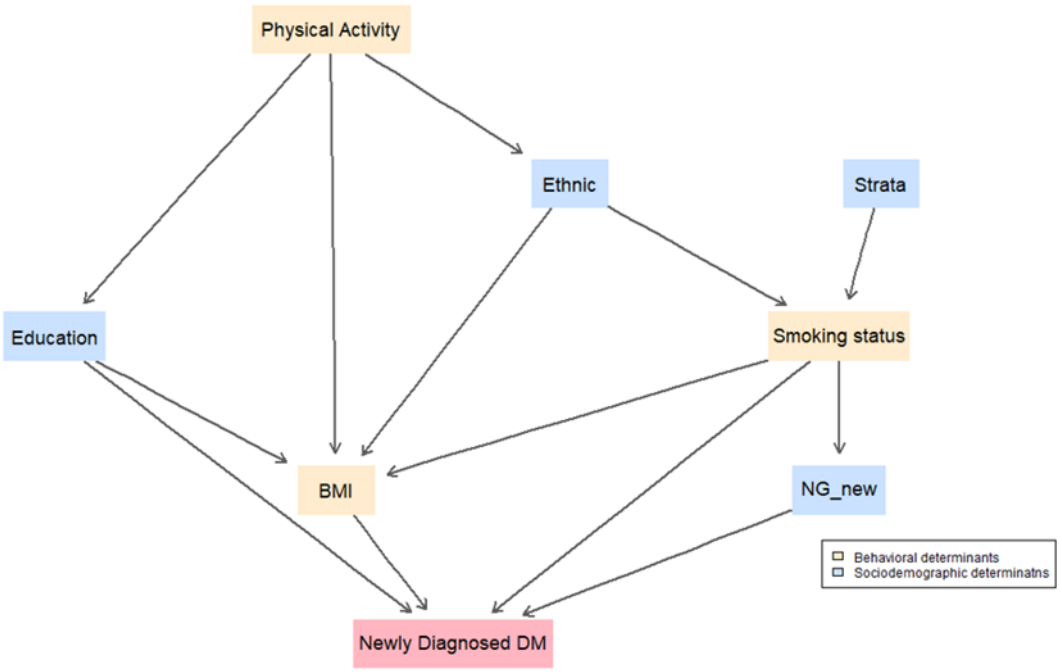

Figure 1: Graphical Network for newly diagnosed Diabetes Mellitus

DISCUSSION

High prevalence of diabetes mellitus among Malaysians have been around for some time as indicated by previous NHMS surveys[4,5]. Current study is one the few studies to investigate the overall relationships of risk factors of newly diagnosed diabetes mellitus among adolescents in Malaysia. Current study have found several risk factors of newly diagnosed diabetes mellitus and the ones that contribute to high conditional probability of having newly diagnosed diabetes mellitus are body mass index (BMI), education, and physical inactivity.

Current study found respondents with abnormal body mass index, such as underweight, overweight, and obese are among those with higher conditional probability of having newly diagnosed diabetes mellitus. Not only do these factors independently increase the probability of having newly diagnosed diabetes mellitus, the interaction between education levels further potentiate the probability[21]. It corresponds to the previous studies carried out in Malaysia, which reported overweight and obesity as significant factors of diabetes mellitus[21-23]. Current study also suggest new findings as underweight respondents were found to have high conditional probability of having newly diagnosed diabetes mellitus.

Education has recognized as one of the preventive action to curb the increment of prevalence of diabetes mellitus, and the relationship between diabetes and education has been found by numerous studies. For example, low education has been identified as a correlate of higher diabetes rates in Canada [15]. Research also found that women with secondary-level education or lower had higher risk compared to those with a tertiary education whereas men with no education had the lowest risk of diabetes [15]. Relationship between education and glycemic control was found to be partially mediated by literacy, especially in an ethnically diverse population with low-income [24]. Low educational levels was found to be associated with incidence of diabetes across different BMI groups and to be used as predictors of DM [25,26]. Our study found that individuals with "unclassified" level of education has the highest prevalence of having newly diagnosed DM followed by no formal education and primary education. This suggests that interventions shall be carried out for both inside and outside of school.

Another concerned finding was the impact of physical inactivity on newly diagnosed DM among Malaysian adolescents and the result of this study has been reinforced because physical activity has a positive effect on type 2 DM [27] .People with highest risk for developing type 2 diabetes are found to be not engaging in regular physical activity and have a sedentary lifestyle, characterized by time spent watching television [28-31] . Therefore, interventions to increase physical activity among Malaysian adolescents are highly required to reduce the risk for developing newly diagnosed diabetes mellitus.

Based on this study, it has been significantly proven that behavioural risk factors which are related to lifestyle, for example overweight and

obesity, physically inactive are associated with newly diagnosed Diabetes Mellitus which will eventually put an impact on the total prevalence of DM because of the progression of NCDs. Prevention is always better than cure. By identifying the modifiable risk factors, lifestyle could be change and improve in order to prevent NCDs. In terms of policy implications, early detection of NCDs are necessary to control and reduce the occurrence and prevalence of newly diagnosed DM . Awareness campaigns or programs could be addressed amongst all Malaysians, particular attention should be paid to the promotion of healthy behaviours, for example, quit smoking and drinking, adequate consumption of fruits and vegetables, maintain normal body weight and being physically active o prevent the development of newly diagnosed DM among Malaysians.

The results of this research are subject to limitations. All behavioral and sociodemographic risk factors relied on self-reporting, which could lead to bias in the results. Secondly, data collected by NHMS 2015 targeted population of age 18 years and above, therefore, results from current study cannot generalize the adolescent population. However, it can generalize to adolescents aged 18-19 years old only which is based on the proper statistical sampling frame of the NHMS2015 study.

For future studies, sedentary behavior should be included as one of the variables in the analysis. Next, it is encouraged to increase the number of attributed variables related to relevant risk factors, for example, family history which will enable us to understand the characteristics of family history of the newly diagnosed DM patients and to predict the likelihood of having the different outcome levels of DM.

CONCLUSIONS

This study is aimed to investigate risk factors of newly diagnosed DM among adolescents in Malaysia. Behavioral and sociodemographic risk factors were found to be significantly associated with the development of newly diagnosed DM among adolescents in this study. Since chronic diseases, for example Diabetes Mellitus placed a substantial economic burden on society which the costs of chronic illness at around three-quarters of the total national health expenditure has been placed by the United States of America (NHMS. 2015). For example, diabetes, account for between 2% and 15% of national health expenditure in some European countries.

As a result, the insightful findings and evidence of this study provides a platform for early detection and prevention of NCDs especially for Newly Diagnosed DM and these needs to be established in order to monitor and control the increasing prevalence of different stages of Diabetes Mellitus in Malaysia. Hence, the prevention of future mortalities and morbidities of different stages of Diabetes Mellitus would effectively reduce the economic burden on the allocation of resources especially on the treatment cost of the DM of the country.Effective interventions are essential to be built upon these information in order to decrease the future economic burden in our country.

ACKNOWLEDGEMENT

The authors also want to thank the General Director of the Ministry of Health in Malaysia for his permission to use the data from the Fifth National Health and Morbidity Survey (NHMS V) and to publish this paper.

Competing Interests
None.

Funding
The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported *by* the Ministry of Higher Education (MOHE) through Fundamental Research Grant Scheme (FRGS/1/2019/SS08/HWUM/02/1).

REFERENCES

1. Centers of Disease Control and Prevention. Diabetes basics. https://www.cdc.gov/diabetes/basics/diabetes.html. Published 2020. Updated 11 June 2020. Accessed 17 July, 2020.
2. National Institute of Diabetes and Digestive and Kidney Diseases. Diabetes overview. https://www.niddk.nih.gov/health-information/diabetes/overview. Published 2016. Accessed 17 July, 2020.
3. World Health Organization. Diabetes. https://www.who.int/news-room/fact-sheets/detail/diabetes. Published 2020. Updated 8 June 2020. Accessed 17 July, 2020.
4. Institute for Public Health. The National Health and Morbidity Survey 2011. Institute for Public Health, National Institutes of Health, Ministry of Health. http://iku.moh.gov.my/images/IKU/Document/REPORT/NHMS2011-VolumeII.pdf. Published 2011. Accessed 11 August, 2020.
5. Institute for Public Health. The National Health and Morbidity Survey 2015- Vol. II: Non-Communicable Diseases, Risk Factors & Other Health Problems. Institute for Public Health, National Institutes of Health, Ministry of Health. http://iku.moh.gov.my/images/IKU/Doc

ument/REPORT/nhmsreport2015vol2.pdf. Published 2015. Accessed 11 August, 2020.

6. Azizi E, Borzouei S, Soltanian AR. A Nonparametric Bootstrap Confidence Interval for the Prevalence of Undiagnosed Type 2 Diabetes Mellitus in Hamadan, West of Iran. *Journal of research in health sciences*. 2018;18(3):e00421.

7. Mehmood K, Junaid N. Prevalence of undiagnosed type 2 diabetes mellitus in Pakistan: Results of screen-diabetes disease registry. *JPMA*. 2018;68(1171).

8. Siu S, Wong K, Lee K, et al. Prevalence of undiagnosed diabetes mellitus and cardiovascular risk factors in Hong Kong professional drivers. *Diabetes research and clinical practice*. 2012;96(1):60-67.

9. Bener A, Zirie M, Janahi IM, Al-Hamaq AO, Musallam M, Wareham NJ. Prevalence of diagnosed and undiagnosed diabetes mellitus and its risk factors in a population-based study of Qatar. *Diabetes research and clinical practice*. 2009;84(1):99-106.

10. Pramono LA, Setiati S, Soewondo P, et al. Prevalence and predictors of undiagnosed diabetes mellitus in Indonesia. *Age*. 2010;46(53.6):100.100.

11. Rathmann W, Haastert B, Icks A, et al. High prevalence of undiagnosed diabetes mellitus in Southern Germany: target populations for efficient screening. The KORA survey 2000. *Diabetologia*. 2003;46(2):182-189.

12. Moy FM, Gan CY, Zaleha MKS. Body mass status of school children and adolescents in Kuala Lumpur, Malaysia. *Asia Pacific journal of clinical nutrition*. 2004;13(4):324-329.

13. Leahy S, O'Halloran A, O'Leary N, et al. Prevalence and correlates of diagnosed and undiagnosed type 2 diabetes mellitus and pre-diabetes in older adults: Findings from the Irish Longitudinal Study on Ageing (TILDA). *Diabetes research and clinical practice*. 2015;110(3):241-249.

14. Cheng YJ, Kanaya AM, Araneta MRG, et al. Prevalence of diabetes by race and ethnicity in the United States, 2011-2016. *Jama*. 2019;322(24):2389-2398.

15. Creatore MI, Moineddin R, Booth G, et al. Age-and sex-related prevalence of diabetes mellitus among immigrants to Ontario, Canada. *Cmaj*. 2010;182(8):781-789.

16. Noor S, Bushara S, Sulaiman A, Elmadhoun WM, Ahmed M. Undiagnosed diabetes mellitus in rural communities in Sudan: prevalence and risk factors. 2015. *Eastern MediterraneanHealth Journal*. 2015; 21(3).

17. Institute for Public Health. The National Health and Morbidity survey 2015 Methodology & General Findings. Institute for Public Health. http://iku.moh.gov.my/images/IKU/Document/REPORT/nhmsreport2015vol1.pdf. Published 2015. Accessed 10 May 2021.

18. AIMultiple. Data Cleaning: What it is, why it matters, best practices & tools. https://research.aimultiple.com/data-cleaning/#what-are-the-benefits-of-data-cleaning. Published 2020. Updated 10 August 2020. Accessed 12 August, 2020.

19. Lewandowski P. What is data cleaning and why is it important? https://sunscrapers.com/blog/why-is-clean-data-so-important-for-analytics-and-business-intelligence/. Published 2018. Accessed 6 August, 2020.

20. Marco Scutari J-BD. *Bayesian Networks with Examples in R*. Chapman and Hall/CRC; 2014.

21. Rahim FF, Abdulrahman SA, Kader Maideen SF, Rashid A. Prevalence and factors associated with prediabetes and diabetes in fishing communities in penang, Malaysia: A cross-sectional study. *PLOS ONE*. 2020;15(2):e0228570.

22. Tee ES, Yap RWK. Type 2 diabetes mellitus in Malaysia: current trends and risk factors. *European Journal of Clinical Nutrition*. 2017;71(7):844-849.

23. Rampal S, Rampal L, Rahmat R, et al. Variation in the Prevalence, Awareness, and Control of Diabetes in a Multiethnic Population: A Nationwide Population Study in Malaysia. *Asia Pacific Journal of Public Health*. 2009;22(2):194-202.

24. Schillinger D, Barton LR, Karter AJ, Wang F, Adler N. Does Literacy Mediate the Relationship between Education and Health Outcomes? A Study of a Low-Income Population with Diabetes. *Public Health Reports*. 2006;121(3):245-254.

25. Agardh EE, Sidorchuk A, Hallqvist J, et al. Burden of type 2 diabetes attributed to lower educational levels in Sweden. *Population Health Metrics*. 2011;9(1):60.

26. Sacerdote C, Ricceri F, Rolandsson O, et al. Lower educational level is a predictor of incident type 2 diabetes in European countries: The EPIC-InterAct study. *International Journal of Epidemiology*. 2012;41(4):1162-1173.

27. Jan Mohamed HJB, Yap RWK, Loy SL, Norris SA, Biesma R, Aagaard-Hansen J. Prevalence and Determinants of Overweight, Obesity, and Type 2 Diabetes Mellitus in Adults in Malaysia. *Asia Pacific Journal of Public Health*. 2014;27(2):123-135.

28. Dunstan DW, Salmon J, Owen N, et al. Physical Activity and Television Viewing in Relation to Risk of Undiagnosed Abnormal Glucose Metabolism in Adults. *Diabetes Care*. 2004;27(11):2603.

29. Morrato EH, Hill JO, Wyatt HR, Ghushchyan V, Sullivan PW. Physical Activity in U.S. Adults With Diabetes and At Risk for Developing Diabetes, 2003. *Diabetes Care*. 2007;30(2):203.

30. Hu FB, Li TY, Colditz GA, Willett WC, Manson JE. Television watching and other sedentary behaviors in relation to risk of obesity and type 2 diabetes mellitus in women. *Jama*. 2003;289(14):1785-1791.

31. Hu FB. Sedentary lifestyle and risk of obesity and type 2 diabetes. *Lipids*. 2003;38(2):103-108.