

# Towards a Parsimonious Pathway Model of Modifiable and Mediating Risk Factors Leading to Diabetes Risk

Yi-Ching Lynn Ho (✉ [yiching.lynn.ho@gmail.com](mailto:yiching.lynn.ho@gmail.com))

Singapore Health Service: Singhealth <https://orcid.org/0000-0001-6317-8525>

**Vivian Shu Yi Lee**

Singapore Health Service: Singhealth

**Moon-Ho Ringo Ho**

Nanyang Technological University

**Gladis Jing Lin**

Singapore Health Service: Singhealth

**Julian Thumboo**

Singapore Health Service: Singhealth

---

## Research

**Keywords:** Diabetes, risk factors, modifiable, mediators, lifestyle, physiological, pathways, mediation analysis, structural equation modeling, Indonesia

**Posted Date:** April 20th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-429757/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

## Abstract

**Background:** The development of diabetes mellitus has been closely linked to multiple risk factors, of which modifiable factors are of particular interest for disease prevention. Yet few studies have assessed the system of pathways through which risk factors lead to diabetes, and how the different groups of risk factors may interact, both as independent or mediating factors.

**Methods:** We aimed to develop a broad pathway model for diabetes risk with modifiable lifestyle risk factors as the start point, hypothesising that Lifestyle Risk (physical inactivity, smoking, poor diet and insufficient sleep) would impact Diabetes Risk (HbA1c) through the mediating factor of Physiological Load (BMI, resting pulse rate, CRP, systolic and diastolic blood pressure). The lifestyle and physiological factors were grouped via principal components analysis and a summary index respectively. Non modifiable risk factors, such as sociodemographics were specified as covariates. We used structural equation modeling to test this model, first using Wave 5 data from the Indonesian Family Life Survey (IFLS), as this was the only wave that collected all indicators of interest. To fit in longitudinal data from an earlier wave (IFLS4), we further tested a smaller model with the two Lifestyle Risk indicators available.

**Results:** Both models showed indirect effects of Lifestyle Risk on Diabetes Risk via Physiological Load, with the cross-sectional model also showing a direct effect. The effect sizes were within the range of other studies that assessed the variables separately.

**Conclusion:** Taken together, the results support the model of an indirect effect of Lifestyle Risk on Diabetes through Physiological Load. Specifying Lifestyle Risk as an observable, composite variable incorporates the cumulative effect of risk behaviour and differentiates this study from previous studies assessing it as a latent construct. We were able to assess causality with retrospective cohort data. Finally, the parsimonious model groups and summarises the multifarious risk factors and illustrates parsimonious and modifiable pathways that could be applied in chronic disease prevention efforts.

## Introduction

Diabetes mellitus alone contributed to 1.6 million deaths worldwide in 2017, a figure that is estimated to double by 2040 (World Health Organization, 2018). In terms of global burden, China (89.5 million), India (67.8 million), USA (30.7 million), Indonesia (21.0 million) and Mexico (13.1 million) have been identified as the top five countries with the most diabetes cases (T2DM accounted for the majority of the cases), together with the most deaths and the highest DALYs due to diabetes (Lin et al., 2020). Among low- and middle income countries, the disease is growing rapidly, with three quarters of diabetics living in these countries (International Diabetes Federation, 2019; World Health Organization, 2016).

The development of T2DM has been closely linked to multiple risk factors, such as an unhealthy diet, physical inactivity, tobacco use, high blood pressure, high body mass index (BMI), sleep and pulse rate, c-reactive protein and sociodemographic factors such as age, gender, ethnicity, and highest level of education attained (Amalia, Cadogan, Prabandari, & Filippidis, 2019; Christiani, Byles, Tavener, & Dugdale, 2015; De Rekeneire et al., 2006; International Diabetes Federation, 2015; Monterrosa, Haffner, Stern, & Hazuda, 1995; Quatromoni, P. A., Copenhafer, D. L., D'Agostino, R. B., & Millen, 2002; World Health Organization, 2011b). These risk factors were established by epidemiological studies exploring them as independent variables in direct relationship with diabetes (Lidfeldt, Nerbrand, Samsioe, & Agardh, 2005; Ng, Sutradhar, Yao, Wodchis, & Rosella, 2020; Persson et al., 2000; Tripathy, Thakur, Jeet, & Jain, 2018). Aside from the sociodemographic and genetic factors, the other risk factors are regarded as modifiable (World Health Organization, 2011b) and can be further differentiated in terms of their temporality. From the start point of a normal baseline, indicators of unfavourable metabolic and physiological changes, such as weight gain and increased blood pressure (World Health Organization, 2011b), can be regarded as downstream factors caused by preceding unhealthy lifestyle-related behaviour (e.g. a lack of physical activity, poor diet, smoking and lack of sleep) (Bardenheier et al., 2013). Obesity is known as the leading risk factor for Type 2 diabetes (Barnes, 2011), but in order to prevent obesity itself, we need to go upstream to the risk factors for high BMI. Indeed physiological risk factors are often associated with lifestyle behavioural factors (Forrester, Leoutsakos, Gallo, Thorpe, & Seeman, 2019; Gay et al., 2015; World Health Organization, 2011a). As a risk factor for diabetes, high BMI is associated with an unhealthy diet that includes high consumption of sodium and saturated fats (Paradis, Godin, Pérusse, & Vohl, 2009; Schulze & Hu, 2005; World Health Organization, 2016). Yet relatively few studies to date have assessed the system of pathways and how the different groups of risk factors may influence each other (both as independent and dependent factors) to impact Diabetes Risk (Bardenheier et al., 2013; Fogelholm, 2010).

In this study, we aimed to develop a broad pathway model for diabetes risk, which would be flexible enough to incorporate different risk factors of interest, as well as serve to outline how diabetes risk may be minimised or prevented from the perspective of health behaviour promotion. It is with this in mind that the modifiable lifestyle risk factors were set as the start point of the pathways. Risk factors that cannot be modified, such as genes or sociodemographics, though determinants, will be specified as covariates, rather than as main factors.

Specifically we aimed to examine the direct and indirect effects of two groups of modifiable risk factors on Diabetes Risk, namely "Lifestyle Risk" and "Physiological Load". On one hand, Lifestyle Risk variables form a natural grouping of independent variables concerned with health-related behaviour, such as the level of physical inactivity, smoking, the consumption frequency of unhealthy food, and insufficient sleep (Amalia et al., 2019; Christiani et al., 2015; International Diabetes Federation, 2015; Monterrosa et al., 1995; Quatromoni, P. A., Copenhafer, D. L., D'Agostino, R. B., & Millen, 2002). On the other hand, the group of Physiological Load variables comprise clinical monitoring indicators measured from individuals, such as body mass index (BMI), resting pulse rate (RPR), c-reactive protein (CRP), systolic (SBP) and diastolic (DBP) blood pressure (Bardenheier et al., 2013; Forrester et al., 2019; Gay et al., 2015; World Health Organization, 2011b, 2011a), and may not be fully independent variables.

We thus hypothesised that Lifestyle Risk would have a positive but indirect effect on Diabetes Risk through the mediator of Physiological Load [Figure 1], as Lifestyle Risk is postulated to be upstream of Physiological Load. Using data from the longitudinal Indonesian Family Life Survey (IFLS), which had administered data collection in five survey waves from 1993 to 2015, we performed structural equation modeling (SEM) on two main models [Figures 3 and 5]. As the hypothesis concerns temporal effects, longitudinal data would be important. However, as only the latest survey wave (IFLS5) had all the indicators of

interest, we first tested a comprehensive model (Model A) with cross-sectional data from IFLS5. This model includes all four Lifestyle Risk factors of interest. We specified a second model (Model B) to allow the testing of longitudinal data (from IFLS4 to IFLS5), but with only two of the Lifestyle Risk factors, as sleep and food intake were not collected in this wave. We account for sociodemographic risk factors as covariates in these analyses.

## Methods

### Study Design and Respondents

In this retrospective cohort study, we used publically available data from the Indonesian Family Life Survey (IFLS), which is an ongoing longitudinal survey that collects socioeconomic and health data in Indonesia. The survey is organised by the RAND Corporation in collaboration with Lembaga Demografi, the University of Indonesia, Center for Population and Policy Studies, the University of Gadjah Mada, the University of California, Los Angeles, and SurveyMETER. To date, there have been five survey waves, from 1993 to 2015. A stratified random sampling scheme on the provinces was adopted and the resulting sample included 13 of 27 provinces in Indonesia, representative of 83% of the population. Households that were randomly selected and participated in the first survey in 1993, were followed up in subsequent waves. Sampling and survey methods have been discussed in detail elsewhere (Strauss, Witoelar, & Sikoki, 2016).

We analysed data from Wave 4 ("IFLS4", conducted from 2007–2008) and Wave 5 ("IFLS5", conducted from 2014–2015), given that the indicators of interest were found in these two waves. Both waves had collected individual-level data on anthropometric measurements, morbidity indicators, healthcare utilization, health behaviours, employment and household expenditure (Strauss et al., 2016), though specific data on food frequency, sleep duration and HbA1c were not available in IFLS4. Figure 2 shows how the final number of samples was derived for the two main models under study (Models A and B: see later section on Mediation Analysis). Respondents included were those who had completed information on sociodemographic factors (age, gender, ethnicity, highest level of education attained); measurements from the health assessment (c-reactive protein (CRP), glycosylated haemoglobin (HbA1c), systolic and diastolic blood pressure, height, weight, resting pulse rate); lifestyle behaviours (engagement in physical activities, number of cigarettes smoked, sleep duration, food frequency (the latter two were used only in Model A)); and employment-related questions involving physical labour (frequency of physical effort and heavy lifting in primary jobs). Respondents with sleep duration up to 14 hours were included. In order to reduce the possible confounding effect of medication in this cross-sectional study, respondents self-reported to be on anti-diabetic or anti-hypertensive medication were excluded.

The final study sample for Model A included 4,000 respondents aged 18 and above (12.9% of total 31,102 adult respondents in IFLS5). For Model B, it included 2,027 respondents aged 18 and above (11.7% of total 17,396 adult respondents present in IFLS4 and 5).

### Sociodemographic Factors

Sociodemographic factors included age, gender, ethnicity and highest level of education attained. Gender was defined as male or female. Ethnicity was grouped into 3 categories: Javanese; Sundanese; and others, similar to other studies (Ananta, Arifin, Hasbullah, Handayani, & Pramono, 2015; Finkelstein, Chay, & Bajpai, 2014). This is because of the 27 distinct ethnic groups sampled in the IFLS, Javanese and Sundanese made up the largest groups. The other ethnicities each made up less than 6% of the sample population, and thus were combined into the Others category (Ananta et al., 2015; Finkelstein et al., 2014). The highest level of education attained was grouped into 4 categories: no education (unschooling); elementary (grade school, kindergarten, Islamic elementary school, adult education A); high school (general junior high, vocational junior high, Islamic Junior high, adult education B, general senior high, vocational senior high, Islamic senior high school, adult education C, Pesantren boarding school); and college/university (open university, Diploma, University) (Amalia et al., 2019; Finkelstein et al., 2014; Hussain, Al Mamun, Reid, & Huxley, 2016).

### Estimation of Lifestyle Risk

Four unhealthy lifestyle behaviours (physical inactivity, smoking, consumption frequency of unhealthy food, and insufficient sleep) were used as indicators of Lifestyle Risk (though in Model B, only physical inactivity and smoking were available for testing).

The level of physical activity was assessed through a modified short form questionnaire in the International physical activity questionnaire (IPAQ) used in IFLS4 and 5. It consisted of a set of self-reported questions for the duration spent on different types of physical activities in the last 7 days. If respondents reported engaging in walking, or any moderate or vigorous physical activities, the total duration of activities will be converted to Metabolic Equivalent of Task (MET)-hours per week using the IPAQ recommended formula to estimate the overall physical activity in a week (Blasio, Donato, & Mazzocco, 2005). The overall physical activity was reverse-coded and used as a measure of "physical inactivity", such that "high physical inactivity" would mean few MET-hours per week and would hence contribute to Lifestyle Risk. The negatively-phrased physical inactivity was used to maintain a consistent interpretation of Lifestyle Risk indicators.

The number of cigarettes smoked was used as an indicator for smoking intensity in this study. This indicator was collected by questionnaires in IFLS4 and 5, which included questions on smoking status, whether respondent has ever smoked self-rolled cigarettes, manufactured cigarettes or cigars, followed by questions on smoking intensity, how many cigarettes/cigars smoked currently or before quitting. Respondents who reported "No" for questions on smoking status would have a smoking intensity of zero.

Unhealthy foods are highly processed food items that are higher in saturated fat, sugar and sodium, such as instant noodles, sweet snacks, fried snacks, fast food or soft drinks (Baker & Friel, 2014). A high consumption of such foods has been shown to exacerbate the risk of chronic diseases (Micha et al., 2017). Therefore, for the purpose of this study, the consumption frequency of unhealthy foods was used as an indicator of Lifestyle Risk. The consumption frequency of unhealthy foods in the past seven days refers to the number of days such foods (instant noodles, sweet snacks, fried snacks, fast food or soft drinks) were consumed in the past week, as captured by the Food Frequency Section of IFLS5. The total consumption frequency of unhealthy food was

calculated by summing the consumption frequencies of individual unhealthy food items with saturation at a score of seven, as the focus was on the consumption frequency of any type of unhealthy food within a week, therefore the range of scores was 0–7.

Sleep duration was determined by two self-reported questions in IFLS5, the time the respondent slept the night before and the time the respondent woke up the day before the survey. The time respondent slept was subtracted by the time respondent woke up to compute sleep duration. Sleep duration was then reverse-coded and used as a measure of insufficient sleep, such that a shorter sleep duration corresponded to more insufficient sleep. The negatively-phrased insufficient sleep was used to maintain a consistent interpretation of Lifestyle Risk indicators.

To specify an appropriate structural equation model, we assessed the suitability of Lifestyle Risk to be a latent construct within the model. A statistical check for an underlying latent construct requires its effect indicators to be positively correlated with one another (Bollen & Bauldry, 2011). Spearman's rank test was performed to test the correlations among the Lifestyle Risk indicators and among the Physiological Load indicators. The correlation results of  $r < |0.4|$  showed that it would not be appropriate to create a latent variable of Lifestyle Risk with their respective indicators as effects. See Additional file 1. (A Spearman's rank correlation coefficient of  $r < |0.4|$  indicates poor correlation.) The Lifestyle Risk indicators would be better identified as a composite variable, because it does not violate this conceptual check (Bollen & Bauldry, 2011). Furthermore, the composite or cumulative effect of these high risk lifestyle behaviours is associated with higher health risk (Sotos-Prieto et al., 2015). As a composite score of high risk lifestyle behaviours, as this means that Lifestyle Risk can be estimated, rather than it being an underlying, unobservable construct.

As there are no standardized thresholds for risk levels in lifestyle behaviours, we used principal component analysis (PCA) to derive the weighted composites of the Lifestyle Risk indicators (physical inactivity, smoking, consumption frequency of unhealthy food, and insufficient sleep). (PCA for Model B only included the first two indicators, as the latter two were unavailable in IFLS4.) The Lifestyle Risk indicators were included as continuous variables in the PCA models, and the principal components (i.e. weighted composites) were orthogonally transformed using Varimax rotation and underwent Kaiser normalization to achieve a structure with independent components for greater interpretability. Principal components with eigenvalue  $> 1$ , scree test and parallel analysis were considered in determining the number of components to retain (Cattell, 1966; Franklin, Gibson, Robertson, Pohlmann, & Fralish, 1995; Kaiser, 1960).

## Estimation of Physiological Load

In this study, Physiological Load is conceptualised as an estimate of the cumulative physiological burden or stress on the body system, similar to the concept of allostatic load or “cumulative biological risk” (McEwen & Stellar, 1993). Allostatic load describes the physiological consequences of the body's attempts to adapt to chronic stressors, which may result in dysregulation spreading among multiple body systems, potentially combining to increase disease risk (McEwen, 1998b; McEwen & Stellar, 1993). Summary measures are typically used to characterise allostatic load across the cardiovascular, metabolic, immune, nervous, and hormonal systems (McEwen, 1998a; McEwen & Stellar, 1993).

In this study, we used five available biomarkers (body mass index, resting pulse rate, c-reactive protein, systolic and diastolic blood pressure) from IFLS4 and 5 to calculate a summary measure of Physiological Load. (Given that this study is not assessing the full list of biomarkers that characterise allostatic load, such as cortisol or catecholamines, which reflect the nervous and hormonal systems (McEwen, 1998a), we have used a separate term “Physiological Load”.) Each of these Physiological Load indicators has been shown to be associated with diabetes (D. Aune, ó Hartaigh, & Vatten, 2015; Bays, Chapman, & Grandy, 2007; De Rekeneire et al., 2006; Hayashi et al., 1999).

The body mass index (BMI) was computed by taking the weight (kilograms) divided by the height (metres) squared. Resting pulse rate (RPR) and blood pressure were averaged from three measurements on alternate arms while respondents were seated upright. CRP concentrations were derived from finger prick dried blood samples. The concentrations were measured by a high-sensitivity CRP (hsCRP) enzyme-linked immunosorbent assay (ELISA) method, which used a hsCRP enzyme immunoassay kit (Herningtyas et al., 2017). CRP plasma equivalent values were used.

Prior to creating a summary measure of Physiological Load, we performed a conceptual check, similar to the one done for Lifestyle Risk, in order to ascertain suitability for specification as a latent construct. We assessed the Spearman correlations of the Physiological Load indicators and found low correlations  $r < |0.4|$  between them (except for SBP and DBP, which are closely related), indicating that the creation of a latent variable using the Physiological Load indicators as effect indicators would not be suitable (Bollen & Bauldry, 2011). [See Additional file 1]. We then proceeded to create a summary index of Physiological Load by using the method similar to allostatic load summary measurements – for each biomarker, a score of one is given for values beyond a clinical threshold reflecting high risk, with a score of zero given if otherwise (Slopen, Non, Williams, Roberts, & Albert, 2014). These are then added up to form a non-weighted summary index (range of 0 to 5) of Physiological Load for each respondent. High risk was defined as: BMI of  $\geq 25$  (Nishida et al., 2004), RPR of  $\geq 90$  bpm (Seccareccia et al., 2001), SBP  $\geq 140$  mmHg and DBP  $\geq 90$  mmHg (Chobanian et al., 2003). CRP  $\geq 30$  mg/L (3 mg/dl) (Ridker, 2003).

## Estimation of Diabetes Risk

As the primary outcome variable, the risk of diabetes was estimated with the level of glycosylated haemoglobin (HbA1c), a surrogate biomarker of average glycemic control over the previous three months. HbA1c was assayed using dried blood samples from the IFLS respondents using a validated protocol (Herningtyas et al., 2017; Hu et al., 2015), and was available only in IFLS5.

## Structural Equation Modeling (Mediation analysis)

Descriptive characteristics of all the variables were generated: categorical variables (e.g. gender, ethnicity) were summarised through counts and percentages, while all the other variables were continuous and were summarised through minimum, maximum, median and interquartile ranges (IQR).

Mediation analysis (Gunzler, Chen, Wu, & Zhang, 2013) via SEM (Hoyle, 2012) was used to test the hypothesis that Lifestyle Risk has a positive but indirect effect on Diabetes Risk through Physiological Load mediators. In the models, weighted composites representing Lifestyle Risk were considered independent

(exogenous) variables, Physiological Load was a mediator and HbA1c was the marker of the outcome variable of Diabetes Risk. Sociodemographic variables were controlled for as covariates of Lifestyle Risk, Physiological Load and HbA1c (Bollen & Bauldry, 2011).

Two main models were tested. Model A was a comprehensive model, which included all the indicators of interest. It would be ideal to use longitudinal data, as the hypothesis concerns temporal effects. However, only IFLS5 had all the indicators of interest. Therefore we first tested the comprehensive model with cross-sectional data from IFLS5. We then created a smaller model (Model B) that allowed us to do a longitudinal analysis, using data from IFLS4. This model was similar to Model A, except for the Lifestyle Risk indicators of the frequency of unhealthy food consumption and sleep duration, both of which were unavailable in IFLS4. Physiological Load was modelled as the mean of Physiological Load from IFLS4 and 5, in order to obtain a value that would be temporally midway between the surveys of IFLS4 and 5. The outcome of Diabetes Risk was represented by the HbA1c values, available only in IFLS5. Figure 3 illustrates Model A, and Fig. 5 illustrates Model B.

The maximum likelihood procedure that provides correction to estimates, standard errors, and a mean- and variance-adjusted chi-square test statistic robust to non-normality was adopted (Byrne, 2012). Standardized estimates with 95% confidence intervals (95% CI) and p-values were reported. To evaluate model fit, the root mean squared error of approximation (RMSEA), comparative fit index (CFI), Tucker-Lewis index (TLI), and standardized root mean squared residual (SRMR) were reported. The following values indicate good model fits: RMSEA value < 0.05, CFI and TLI > 0.95, and SRMR < 0.08. (West, Taylor, & Wu, 2012).

To corroborate the results of Model A, we compared it with an alternative outcome measurement (Model A2) by replacing the continuous outcome variable of Diabetes Risk with a binary variable: using the World Health Organisation (WHO)-recommended threshold of HbA1c  $\geq$  6.50% for diabetes diagnosis (World Health Organization, 2011c), respondents were categorised accordingly.

Data preparation and regression analyses were performed using STATA (version 14.0) software (StataCorp, 2015). PCA and parallel analysis was performed on Rstudio (R version 4.0.2) software, using *psych* package (version 1.9.12) and *paran* package (version 1.5.2), respectively (Dinno, 2018; Revelle, 2019; RStudio Team, 2020). All mediation analyses were performed using Mplus 8 (version 1.5) software (Muthén & Muthén, 2017). In all statistical analyses,  $p < 0.05$  and 95% CI that do not include zero were considered statistically significant.

## Post-hoc analyses

To investigate the possibility that the involvement of physical labour in respondents' primary jobs confounded physical inactivity levels of Lifestyle Risk composite variables, physical labour was assessed by responses to two self-reported questions, whether respondents' primary jobs involved physical effort or heavy lifting all the time, almost all the time, or most of the time. In this model (Model A3), involvement of physical labour in primary jobs was classified as "yes" or "no" based on their responses, and adjusted for as a covariate of Lifestyle Risk composite variables.

To mitigate the weakness that the analysis for Model A used cross-sectional data, we compared the model fit results against an alternative model with reversed directional relationships (Model A4). The model selection index, Bayesian information criterion (BIC), was used to compare Model A against Model A4, with the smaller BIC value indicating a better model fit.

## Results

### Sample characteristics and comparisons with the national population

Table 1 presents a descriptive summary of demographic characteristics of the samples used in the analysis of Model A and Model B. It also includes summary statistics of the Lifestyle Risk indicators, Physiological Load indicators, and HbA1c levels.

Table 1  
Socio-demographic characteristics, lifestyle and physiological characteristics.

Characteristics	IFLS5 sample					IFLS4 sample						
	n(%) or Median(IQR)			Minimum	Maximum	Proportion of sample beyond clinical threshold	n(%) or Median(IQR)			Minimum	Maximum	Proportion of sample beyond clinical threshold
Age, median(IQR)	40.00	(31.0,	56.0)	18.0	102	-	55.00	(49.0,	62.0)	23.0	88.0	-
<b>Gender, n (%)</b>												
Male	2078	(53.0)		-	-	-	879	(43.4)		-	-	-
Female	1854	(47.0)		-	-	-	1148	(56.6)		-	-	-
<b>Ethnicity, n (%)</b>												
Javanese	1853	(47.6)		-	-	-	982	(48.4)		-	-	-
Sundanese	425	(10.8)		-	-	-	228	(11.2)		-	-	-
Others	1654	(41.6)		-	-	-	817	(40.3)		-	-	-
<b>Highest Education level, n (%)</b>												
No education	235	(5.93)		-	-	-	360	(17.8)		-	-	-
Elementary	1360	(34.6)		-	-	-	1139	(56.2)		-	-	-
High school	1810	(46.1)		-	-	-	424	(20.9)		-	-	-
College/University	527	(13.4)		-	-	-	104	(5.13)		-	-	-
<b>Lifestyle Risk indicators, median(IQR)</b>												
Physical activity levels (MET h/w)	31.54	(8.25,	86.7)	0.00	428	-	48.96	(17.0,	113)	0.00	428	-
Smoking (number of cigarettes smoked in a day)	0.00	(0.00,	10.0)	0.00	60.0	-	0.00	(0.00,	6.00)	0.00	80.0	-
Consumption frequency of unhealthy food (days in a week)	5.00	(2.00,	7.00)	0.00	7.00	-	-	-	-	-	-	-
Sleep duration (hr)	7.00	(5.67,	8.00)	0.250	14.0	-	-	-	-	-	-	-
<b>Physiological Load indicators, median(IQR)</b>												
BMI (kg/m <sup>2</sup> )	22.56	(19.9,	25.9)	12.8	51.4	30.7%	22.12	(19.7,	25.4)	11.4	97.2	26.9%
Resting pulse rate (bpm)	74.67	(68.0,	82.3)	39.0	139	9.58%	75.33	(69.3,	82.0)	45.2	120	10.2%
CRP (mg/L)	0.75	(0.263,	2.05)	0.00100	91.5	0.375%	0.84	(0.350,	1.99)	0.101	43.4	0.00%
Systolic blood pressure (mmHg)	125.67	(116,	138)	89.3	245	22.9%	136.00	(124,	152)	94.2	227	34.9%

Descriptive are presented in n(%) for categorical variables, median (interquartile range), minimum and maximum for continuous variables. All values were rounded off to 3 significant figures.

1. Age was calculated based on respondents' age in year 2015 for IFLS5 and respondents' age in year 2008 for IFLS4.

2. Physical activity and Sleep duration were later reverse-coded by multiplying with (-1) to maintain a consistent negative interpretation of all Lifestyle Risk indicators.

3. The clinical thresholds for Physiological Load indicators were BMI  $\geq 25\text{kg/m}^2$ ; resting pulse rate  $\geq 90\text{bpm}$ ; c-reactive protein(CRP)  $\geq 30\text{mg/L}$ ; systolic blood pressure  $\geq 140\text{mmHg}$ ; diastolic blood pressure  $\geq 90\text{mmHg}$ ; HbA1c  $\geq 6.50\%$ .

	IFLS5 sample						IFLS4 sample					
Diastolic blood pressure (mmHg)	78.00	(71.7,	85.7)	39.0	142	14.9%	80.17	(73.8,	87.5)	40.0	131	18.8%
<b>Diabetes Risk, median(IQR)</b>												
HbA1c level (%)	5.45	(5.10,	5.85)	3.50	14.0	6.83%	-	-	-	-	-	-
<i>Descriptive are presented in n(%) for categorical variables, median (interquartile range), minimum and maximum for continuous variables. All values were rounded off to 3 significant figures.</i>												
1. Age was calculated based on respondents' age in year 2015 for IFLS5 and respondents' age in year 2008 for IFLS4.												
2. Physical activity and Sleep duration were later reverse-coded by multiplying with (-1) to maintain a consistent negative interpretation of all Lifestyle Risk indicators.												
3. The clinical thresholds for Physiological Load indicators were BMI $\geq 25\text{kg/m}^2$ ; resting pulse rate $\geq 90\text{bpm}$ ; c-reactive protein(CRP) $\geq 30\text{mg/L}$ ; systolic blood pressure $\geq 140\text{mmHg}$ ; diastolic blood pressure $\geq 90\text{mmHg}$ ; HbA1c $\geq 6.50\%$ .												

The IFLS5 sample was similar in composition to the Indonesian general population, with some minor differences. The median age was 40 years (range 18–102 years) with 53.0% males. 59.5% of the respondents attained highest education level of high school and above (High school = 46.1%; College/university = 13.4%). Our IFLS5 sample characteristics were representative of the Indonesian population, where 50.3% of the national population were males, the two largest ethnic groups were also Javanese and Sundanese, and 54.9% of the population attained an education level of at least high school and above (Ananta et al., 2015; Statistics Indonesia - Badan Pusat Statistik - BPS, National Population and Family Planning Board - BKKBN/Indonesia, Kementerian Kesehatan - Kemenkes - Ministry of Health/Indonesia, 2013; Suryadinata, Arifin, & Ananta, 2003; United Nations Statistics Division, 2015).

The median age for the IFLS4 sample was 55 years (range 23–88 years) with 43.4% males. The Javanese and Sundanese were the two largest ethnic groups, similar to Indonesia statistics (Ananta et al., 2015; Statistics Indonesia - Badan Pusat Statistik - BPS, National Population and Family Planning Board - BKKBN/Indonesia, Kementerian Kesehatan - Kemenkes - Ministry of Health/Indonesia, 2013; Suryadinata et al., 2003). In terms of the highest level of education, the largest proportion found in IFLS4 was elementary school (56.2%). Differences in these factors between the IFLS4 sample and the national statistics can be attributed to the filtering of respondents that did not meet the inclusion criteria as outlined in the Methods [see Fig. 2]. Differences between the IFLS4 and IFLS5 samples could also be attributed to change in demographic progression across the 7–8 years between Waves 4 and 5, such as the increase in the proportion of those with higher education.

Figure 4 illustrates the distribution of the HbA1c level among the respondents. The median HbA1c level was 5.45% (range 3.50–14.0%). Applying the WHO cut-off for indication of diabetes, i.e. HbA1c  $\geq 6.50\%$  (World Health Organization, 2011c), 6.83% of our respondents were diabetic [see Table 1], which is very similar to the 6.90% national diabetes prevalence reported in 2013 Indonesia Basic Health Research survey (RISKESDAS) (The National Institute of Health Research and Development, 2013), indicating representativeness of our sample.

## Correlations among Lifestyle Risk and Physiological Load indicators

Relatively low Spearman's rank correlation coefficients of  $r < |0.4|$  were observed between Lifestyle Risk indicators [see Additional file 1]. Similarly, low correlations were found for Physiological Load indicators, with the exception of SBP and DBP, which were to be expected. These results did not support the specification of both Lifestyle Risk and Physiological Load as effect indicators of latent variables (Bollen & Bauldry, 2011). Lifestyle Risk was therefore specified as a composite variable, while Physiological Load was specified as a summary index, similar to allostatic load indices.

## Principal Components of Lifestyle Risk

Following PCA to determine the composite weights of Lifestyle Risk, we selected the first two components for Lifestyle Risk data in IFLS5 (Model A) upon inspection of the scree plot, applying the Kaiser-Guttman rule, and parallel analysis. For IFLS4 Lifestyle Risk data (Model B), we also selected the first two components, as they passed the Kaiser-Guttman rule, though parallel analysis recommended retaining only the first component. The positive loadings of the second component were also in line with our hypothesis. The two components in both models were termed, *LR1* and *LR2*. The total proportion of variance explained by the selected components was 54.9% for IFLS5 data and 100% for IFLS4 data.

Table 2 presents the retained components and their loadings. For IFLS5, the strong loadings in LR1 were physical inactivity (-0.75) and smoking (0.74), while the strong loadings in LR2 were consumption frequency of unhealthy food (0.72) and insufficient sleep (0.71). All component loadings were positive, except for physical inactivity (-0.75) and the consumption frequency of unhealthy food (-0.10) in LR1. For the IFLS4 dataset, the loadings of LR1 were similar to that of IFLS5, where the loadings for physical inactivity and smoking were -0.75 and 0.75 respectively. The loadings of LR2 were 0.66 for both physical inactivity and smoking.

Table 2  
Component matrix of the 2-component principal component analysis (PCA).

Lifestyle risk indicators (IFLS5)	LR1	LR2
Physical inactivity	-0.75	0.12
Smoking	0.74	0.14
Consumption frequency of unhealthy food	-0.10	0.72
Insufficient sleep	0.13	0.71
Lifestyle risk indicators (IFLS4)	LR1	LR2
Physical inactivity	-0.75	0.66
Smoking	0.75	0.66
<i>Principal components were orthogonality transformed using varimax rotation. All values were rounded off to 2 decimal places.</i>		
1. In PCA of Lifestyle Risk indicators in IFLS5, physical inactivity, consumption frequency of unhealthy food, insufficient sleep and smoking were included. Proportion of variance for the first component(LR1) was 28.27%; Proportion of variance for the second component(LR2) was 26.63%.		
2. In PCA of Lifestyle Risk indicators in IFLS4, physical inactivity and smoking were included. Proportion of variance for the first component(LR1) was 55.97%; Proportion of variance for the second component(LR2) was 44.03%.		

## Mediation analyses: The indirect effects of Lifestyle Risk on Diabetes Risk through Physiological Load

Model A: Table 3 presents the direct and indirect relationships between Lifestyle Risk and Diabetes Risk (HbA1c). The model had good fits (RMSEA < 0.05; CFI and TLI > 0.95; SRMR < 0.08). For the comprehensive SEM model results including statistics for the sociodemographic covariates, refer to the Additional file 2. Both components of Lifestyle Risk were shown to have effects on Diabetes Risk through the mediator of Physiological Load (PL). LR1 had a significant indirect effect on HbA1c, though it was through an inverse relationship with PL. On the other hand, LR2 had a positive, direct effect on HbA1c, as well as a positive, indirect effect on HbA1c through Physiological Load.

Table 3  
Model A and Model B with standardized estimates, p-values and 95% confidence interval.

	Model A (IFLS5 data)			Model B (IFLS4 & 5 data)		
	Standardized estimate	p-value	95% CI	Standardized estimate	p-value	95% CI
<b>Direct effects on HbA1c</b>						
LR1	-0.012	0.461	(-0.045, 0.020)	0.007	0.817	(-0.054, 0.069)
LR2	<b>0.033</b>	< 0.050	(0.002, 0.064)	0.032	0.244	(-0.022, 0.086)
<b>Indirect effects on HbA1c via Physiological Load mediator</b>						
LR1	<b>-0.011</b>	< 0.050	(-0.019, -0.004)	-0.003	0.633	(-0.013, 0.008)
LR2	<b>0.009</b>	< 0.050	(0.003, 0.016)	<b>0.010</b>	< 0.050	(0.001, 0.020)
Model fit indices	RMSEA	0.014		0.069		
	CFI	1.000		0.988		
	TLI	0.989		0.601		
	SRMR	0.002		0.016		
<i>Significant estimates at p &lt; 0.05 are shown in bold. All values were rounded off to 3 decimal places.</i>						
1. Model A (using only IFLS5 data) adjusted for socio-demographic covariates, age, gender, ethnicity, and highest education level attained in IFLS5 (full direct and indirect-effects model results in Additional file 2).						
2. Model B (using IFLS4 and 5 data) adjusted for socio-demographic covariates, age, gender, ethnicity, and highest education level attained in IFLS4 (full direct and indirect-effects model results in Additional file 6).						
3. Statistically significant estimates, in bold, indicate Physiological Load index mediate the indirect effects of Lifestyle Risk (LR1 and LR2) on HbA1c. Model fit indices were generally within the acceptable threshold and suggested good model fits.						

Model B: Table 3 presents the direct and indirect relationships between Lifestyle Risk and Diabetes Risk (HbA1c). The model had acceptable fits (RMSEA < 0.08; CFI > 0.95; SRMR < 0.08). For the comprehensive SEM model results including statistics for the sociodemographic covariates, refer to the Additional file 3. LR1 did not have a significant effect on Physiological Load or Diabetes Risk. However, LR2 had a positive indirect effect on Diabetes Risk through positive effects on Physiological Load.

## Comparisons of Model A and alternative models

Model A2 (where we used binary outcome for Diabetes (yes if HbA1c  $\geq$  6.50%), instead of continuous HbA1c values): This yielded similar significant indirect effects from both LR1 and LR2 on Diabetes Risk as the hypothesized Model A. Also similar to Model A, there were no direct effect from LR1 on Diabetes Risk. [see Additional file 4]. This corroborated Model A.

With reference to physical activity guidelines (Kaminsky & Montoye, 2014; World Health Organization, 2020), at least 8.30 MET hours / week of physical activity is recommended. Compared to this, the median for the respondents in IFLS4 and IFLS5 was approximately 6 times (49.0 MET hours/week) and 4 times (31.5 MET hours/week) higher respectively [Table 1]. Due to the surprisingly robust physical activity levels, we postulated that these levels could be confounded by jobs that involved physical labour. Indeed, individuals with physical labour in their primary jobs had significantly higher MET levels than individuals who did not,  $t(3998) = 13.1, p < 0.001$ , indicating that the jobs with manual labour contributed to this "lifestyle behaviour". Therefore, involvement of physical labour in respondents' primary jobs was adjusted for as a covariate of Lifestyle Risk composite variables in a post-hoc analysis (Model A3). There was a positive and significant association between involvement of physical labour and the Lifestyle Risk composite variable of LR1 (0.123,  $p < 0.001$ ), but not for LR2. Otherwise, the standardized estimates and significance of direct and indirect relationships in the model were similar to Model A, with model fit indices indicating good fit (RMSEA  $< 0.05$ ; CFI  $> 0.95$ ; TLI  $> 0.80$ ; SRMR  $< 0.08$ ) [see Additional file 5].

Given that Model A was based on cross-sectional data, we compared it to a model in which the pathways were reversed (Model A4). BIC values were lower for Model A than Model A4, indicating better fit [see Additional file 6].

## Sociodemographic covariates

Sociodemographic variables were not the focus of this study, but they are nonetheless important to factor into the models in order to control for the relationships they have with the dependent variables. The significant covariates varied slightly for Model A and Model B, though the result directions were consistent, except for the significance of the "Others" ethnic group factor in its association with Lifestyle Risk.

In summary, age was a factor in Lifestyle Risk (only Model B), Physiological Load and Diabetes Risk (only Model A), with increasing age leading to increasing risk. Gender was a factor in Lifestyle Risk, Physiological Load and Diabetes Risk (only Model A), with females associated with physical inactivity and higher Physiological Load, while males associated with poorer diet and sleep, and Diabetes Risk. Ethnicity was a factor in Lifestyle Risk, Physiological Load and Diabetes Risk (only Model B): the Sundanese were associated with higher Lifestyle Risk, while the results were mixed for the Others ethnic category; the Others were associated with less Physiological Load, while the Sundanese associated with less Diabetes Risk. Finally, education levels were a factor in Lifestyle Risk and Physiological Load (only Model B), with higher education being associated with higher Lifestyle Risk as well as higher Physiological Load.

## Discussion

We have developed a general pathway model from the start point of modifiable lifestyle behaviour, and have demonstrated how the behavioural components in Lifestyle Risk can affect Diabetes Risk via the mediating factor of Physiological Load. The overall results support our hypothesis that the effect of Lifestyle Risk on Diabetes Risk is likely to be indirect, and thus offers a stepwise perspective, whereby upstream and downstream modifiable factors could be modeled pathwaywise. Specifying Lifestyle Risk as an observable composite variable incorporates the cumulative effect of risk behaviour and differentiates this study from previous studies looking at it as a latent construct (Bardenheier et al., 2013; Roman-Urrestarazu et al., 2016; Tripathy et al., 2017). As a retrospective cohort study, it also has the advantage of being able to assess causality (e.g. the longitudinal analyses of Model B), rather than just associations between factors. From a disease prevention perspective, it helps to narrow our focus to an initial set of lifestyle risk factors, from which to monitor the progression of health risk towards more downstream physiological factors to disease.

The use of mediation analysis via SEM allowed us to simultaneously assess multiple pathways within a single model. One of the strengths of SEM is the ability to accommodate a variable to be both independent and dependent (i.e. a mediator) and to simultaneously assess multiple pathways on dependent variables. This differs from traditional regression, one of the most common techniques used in determining the contribution of risk factors to disease outcome. Regression is able to accommodate variables to be either independent or dependent, but not both at the same time. Multiple pathways in a single model need to be tested separately, resulting in a potential multiple comparisons problem (Miller, 1981). As with any technique, limitations exist, such as the theoretical assumption that all directional pathways in the model are known (Pearl, 2012). Nonetheless, the recognition of the usefulness of SEM for analysing pathways in disease is growing (Camelo et al., 2014; Roman-Urrestarazu et al., 2016; Seixas et al., 2018; Tripathy et al., 2017). In diabetes, the interplay of risk factors is complex and SEM may provide a better tool than regression to analyse these relationships.

Only a few studies have examined multiple risk factors simultaneously, as traditionally studies on risk factors in chronic disease mainly used regression analyses that examined the factors independently. The first study that examined all these risk factors simultaneously as a multiple pathway system to diabetes risk was Bardenheier et al, 2013 (Bardenheier et al., 2013). They used 10 variables with 27 hypothesised pathways in an SEM and found through best-fit iterations of the model that physical activity and poor diet were significant lifestyle factors (other lifestyle factors were not studied) that contributed to diabetes risk via large waist circumference, high blood pressure, triglycerides, and HDL. Subsequently several other studies applied similar models in their own population data (Roman-Urrestarazu et al., 2016; Tripathy et al., 2017) with varying results, but generally finding that physical activity and poor diet impact diabetes risk through separate mediators like BMI, blood pressure, high-density lipoprotein (HDL) and triglycerides.

These studies created latent constructs for lifestyle factors, for example, physical activity was modeled as a latent construct, with the effect indicators being house / yard work, muscle strengthening, walking / biking, moderate exercise, and vigorous exercise. A latent variable is an unobservable construct, and it is a determinant or cause of the observed measures (which are known as effect indicators) (Bollen & Bauldry, 2011). Effect indicators should thus be highly correlated with each other, since a shift in the underlying construct would similarly affect all the effect indicators (Bollen & Bauldry, 2011). Therefore in our opinion, a behaviour like physical activity may not be suitable to be modeled as a latent construct, since it is actually observable. Furthermore, the effect

indicators as outlined above, may not actually correlate with each other, given possibly differing motivations for each type of physical activity. In our study, physical activity is treated as an observed indicator resulting from total MET in a week from various levels of physical activity (and is then inverted to represent the lifestyle risk of physical inactivity). Including the other lifestyle risk behaviour (including poor diet, lack of sleep and smoking), we postulated that all these behaviours would *comprise or contribute to* a total composite variable of Lifestyle Risk. Composite indicators can be formed from linear combinations of variables that are uncorrelated and not conceptually similar (Bollen & Bauldry, 2011).

Indeed, from a statistical view point, the very low correlations (from 0.000 to 0.067) that we found between the Lifestyle Risk indicators supported our opinion that they should not be modeled as latent variables in our study. Thus unlike the earlier studies, we modeled these indicators instead as weighted composites of the variable, Lifestyle Risk, as this would allow the useful analysis of the combined effect of the lifestyle behaviours on health states.

## Components of Lifestyle Risk

In both Model A and Model B, two Lifestyle Risk components (LR1 and LR2) were derived from the data as complementary composites of unhealthy behaviour. In both models, LR2 is the simpler and more straightforward of the two Lifestyle Risk components – it encompasses a clear set of unhealthy lifestyle behaviours, with all behaviour indicators having positive coefficients (i.e. the increase of such behaviours represent higher Lifestyle Risk). In Model A, unhealthy diet and lack of sleep were the dominant parts of the component. The other two lifestyle factors of physical inactivity and smoking showed much smaller factor loadings, and therefore were considered minor contributors to Model A's LR2. In Model B, both physical inactivity and smoking contribute equally strongly to LR2.

As hypothesised, an increase in this lifestyle component goes on to increase Diabetes Risk through the mediator of Physiological Load. In Model A, there was also a direct effect of LR2 on Diabetes Risk, although we also bear in mind its cross-sectional nature. Physical activity and poor diet are both well established risk factors corresponding to risk of chronic disease, in particular diabetes (Dagfinn Aune, Norat, Leitzmann, Tonstad, & Vatten, 2015; Bellou, Belbasis, Tzoulaki, & Evangelou, 2018; Mills, Stefanescu, & He, 2020). Smoking is also another risk factor, showing a dose-response phenomenon with risk of diabetes (Willi, Bodenmann, Ghali, Faris, & Cornuz, 2007), while short sleep duration is gaining attention as a factor involved in developing risk of diabetes through associations with BMI and blood pressure (Buxton & Marcelli, 2010; Jarrin et al., 2018; Reutrakul & Van Cauter, 2018). The path coefficients from Lifestyle Risk to Physiological Load are relatively small (0.04–0.05), though the effect sizes are within the range of a study that used regressions to assess lifestyle variables and BMI with the same IFLS dataset (Pengpid & Peltzer, 2017).

No direct effect of Lifestyle Risk was found in the longitudinal Model B, highlighting that caution has to be applied when assessing Lifestyle Risk as a direct risk factor contributing to Diabetes Risk, as it could be further upstream compared to Physiological Load. This aligns with studies showing the indirect effects of physical activity on the risk of developing diabetes through variables like BMI, but without detected direct effects (Bardenheier et al., 2013; Miller, 1981; Seixas et al., 2018).

Compared to LR2, LR1 is an intriguing mix of lifestyle risk behaviour in both models: the two dominant contributors to the component are physical inactivity and smoking. However the composite weightings are in opposite directions – for physical inactivity, it is negative, meaning being physically active, while smoking has the expected positive weighting that contributes to Lifestyle Risk. LR1 can therefore be seen as comprising of both healthy (physical activity) and unhealthy behaviour (smoking). This apparent contradiction in lifestyle may be explained by our post hoc analysis showing that a significant contributor to the high MET scores (as an indicator of physical activity) in LR1 is involvement in physical labour. Thus the “healthy behaviour” of high physical activity in LR1 is likely to be due to work obligations, rather than a voluntary healthy lifestyle per se.

Nonetheless, as a component with mixed effects of lifestyle risk, the results show that an increase in LR1 indirectly reduces Diabetes Risk through a reduction in Physiological Load. This protective effect of LR1 on both Physiological Load and Diabetes Risk may be explained in several ways. First, there is strong evidence that being physically active in various ways reduces BMI (Fogelholm & Kukkonen-Harjula, 2000), RPR (Reimers, Knapp, & Reimers, 2018), blood pressure (Bakker, Sui, Brellenthin, & Lee, 2018) and CRP (Kasapis & Thompson, 2005). Thus it may be possible that in this sample, the extent of physical activity resulted in physiological benefits large enough to counter the negative effects of smoking. Second, while smoking is a well accepted risk factor (World Health Organization, 2011b), it has also been shown to have mixed effects on health indicators, e.g. blood pressure (Sotos-Prieto et al., 2015). Negative associations have been found for smoking and blood pressure (Sotos-Prieto et al., 2015), although when adjusted for life-course, the number of cigarettes smoked correlated with the risk of high blood pressure, indicating that smoking may require time to result in downstream blood pressure effects (Sotos-Prieto et al., 2015). This may perhaps explain the results for Model B, where there was no significant effect from the LR1 component to Physiological Load and Diabetes Risk, as the longitudinal analysis (compared to the cross sectional analysis in Model A) may have allowed time for the downstream negative effects of smoking to counterbalance the protective effects of physical activity.

Taken together, these results show that Lifestyle Risk can significantly impact (and specifically for the LR2 behaviours, to significantly worsen) Physiological Load, which will in turn also influence the risk of diabetes.

## Physiological Load as a Mediator of Diabetes Risk

Physiological Load as a summary index, was found to be a significant mediator in both main models, with every one unit increase in Physiological Load corresponding to an increase in HbA1c value by approximately 0.2 percentage points (based on the unstandardised estimates). The one unit increase in Physiological Load can be achieved by a cross in clinical threshold for any one of the five indicators used to define Physiological Load. The pathway coefficients are also within the ranges that have been found in other studies on pathways to diabetes risk (Bardenheier et al., 2013; Roman-Urrestarazu et al., 2016; Tripathy et al., 2017). The indicators used in determining Physiological Load are routine clinical monitoring indicators and each of them have been associated with risk of diabetes (D. Aune et al., 2015; Gress, Nieto, Shahar, Wofford, & Brancati, 2000; Kahn, Hull, & Utzschneider, 2006; Wang et al., 2013). While they are considered modifiable risk factors of diabetes, they are themselves associated effects of lifestyle behaviour (Bakker et al., 2018; Fogelholm &

Kukkonen-Harjula, 2000; Gay et al., 2015; Kasapis & Thompson, 2005; Paradis et al., 2009; Persson et al., 2000; Reimers et al., 2018), and thus may be considered as intermediate markers of diabetes risk.

Modeling a grouped measure of these Physiological Load indicators was considered, because there is growing research showing firstly, that having adverse risk factors across multiple physiological systems strongly predicts morbidity and mortality (Juster, McEwen, & Lupien, 2010), secondly, that physiological dysfunction can spread across multiple physiological systems and combine to elevate disease risk (McEwen & Seeman, 1999; Slopen et al., 2014), and thirdly, that the cumulative total of physiological dysregulation across such indicators can predict morbidity and mortality risks better than individual indicators (Poulter, 2003; Seeman, Epel, Gruenewald, Karlamangla, & McEwen, 2010; Slopen et al., 2014).

We evaluated that it was not suitable to specify Physiological Load as a latent factor within the model, given the relatively low correlations observed between the Physiological Load indicators (Bollen & Bauldry, 2011). We thus employed a summary measure similar to that of “allostatic load” or “cumulative biological risk”, which refer to summary measures that characterize functioning across the cardiovascular, metabolic, immune, nervous, and hormonal systems, indicating the physiological consequences to chronic stressors on the body (McEwen, 1998b, 1998a; McEwen & Stellar, 1993). The concept derives from the definition of allostasis as the ability of the human body to achieve homeostasis and remain functional through continual adjustments of the internal physiological system (McEwen & Wingfield, 2003). Indeed studies have found that physical activity is associated with lower allostatic load (Forrester et al., 2019; Gay et al., 2015; Petrovic et al., 2016), whereas poor sleep quality (Bei, Seeman, Carroll, & Wiley, 2017; Clark et al., 2014), an unhealthy diet (Mattei, Noel, & Tucker, 2011; van Draanen, Prelip, & Upchurch, 2018) and smoking (Sotos-Prieto et al., 2015) are associated with high allostatic load levels. Patients with T2DM have been shown to have higher allostatic load (Carlsson, Nixon Andreasson, & Wändell, 2011; Steptoe et al., 2014), which is also correlated with higher glycated haemoglobin (Carlsson et al., 2011).

In this study, five of the ten indicators normally used in estimating allostatic load (Seeman et al., 2010) were included, as the other indicators were not available. Nonetheless, the focus was on using typical clinical routine monitoring variables in the metabolic and cardiovascular domains as risk factors of Diabetes Risk, and not on indicators of nervous and hormonal responses to chronic stress (which would be important in studies of chronic stress). As such, the term Physiological Load was used to term our summary measure and avoid confusion with allostatic load.

Given the significant role of Physiological Load as a mediator of Lifestyle Risk with Diabetes Risk that we found, including intermediate markers in such models will help to outline the expected proximal changes to be tracked in the community/clinics, especially in mid to long term public health programs in tackling chronic disease through lifestyle recommendations.

## Sociodemographic Covariates

Sociodemographic factors, such as age, gender, ethnicity, and education level, are important determinants in chronic disease and should be accounted for (Di Milia, Vandelanotte, & Duncan, 2013; Heo et al., 2011). However as they are not modifiable in general, we modeled them as covariates, rather than main factors, in order to focus on the effects of modifiable lifestyle behaviour and physiological load on the risk of diabetes. Given that sociodemographic factors may modulate the main factors of Lifestyle Risk, Physiological Load and Diabetes Risk in different ways, covariates were included to explain each of these main factors [Figures 3 and 5, Additional file 2 and Additional file 3].

In this large sample, 6.83% of respondents had HbA1c values of 6.50% and above and can be classified diabetic, according to WHO criteria (World Health Organization, 2011c). This statistic is very similar to the 6.90% found in a national health survey in Indonesia done around the same time (The National Institute of Health Research and Development, 2013), indicating representativeness of our sample. In this study, we observed that 90% of respondents with HbA1c values of  $\geq 6.50\%$  did not report themselves as having received a diagnosis of diabetes [see Additional file 7]. The incongruence of high HbA1c values with low incidence of awareness / reported diagnosis is a concern for public health efforts. The results of this model (Model A2) using binary Diabetes outcome (HbA1c  $\geq 6.50\%$ , or not) corroborate the results of Model A, which uses continuous HbA1c values. As the binary groups were unbalanced, (i.e. the proportion of the group with HbA1c  $\geq 6.50\%$  is much smaller (6.83%) than the group with lower HbA1c (93.2%)), we focused on using continuous outcomes in the statistical analyses.

## Applications and Limitations

As the purpose of the general model is to provide a simple pathway framework that groups distal (lifestyle) and proximal (physiological) factors, it can technically be applied to any analysis concerned with evaluating the relationship of lifestyle risks with chronic disease via the mediator of physiological risk. Specific composites of the distal or proximal factors can be determined by knowledge of the disease etiology, or driven by the dataset at hand, such as was done with the Lifestyle Risk indicators in this study.

Lifestyle indicators that were important to our model, such as food frequency and sleep duration, were only found in Wave 5 of the IFLS, hence the comprehensive model (Model A) was constrained to using the available cross-sectional data, limiting the causal attribution of Diabetes Risk to Lifestyle Risk via Physiological Load. We assumed, similar to other studies (e.g. Bardenheier et al, 2013(Bardenheier et al., 2013)), that Lifestyle Risk behaviour is relatively habitual and has thus preceded Physiological Load and Diabetes Risk. We mitigated the issue of cross-sectional data analysis in three ways. First, we excluded all the respondents who reported that they were taking medication for diabetes and hypertension, in order to avoid the confounding effect of medication on the Physiological Load indicators. Second, we generated a model with all pathways reversed to compare with our hypothesised model, in order to check that our model with pathways from Lifestyle Risk to Diabetes Risk via Physiological Load was the better fit for the variables of interest, which was shown to be the case. Third, we did longitudinal analysis using a smaller model (Model B), without the Lifestyle Risk indicators of food and sleep. This longitudinal analysis supported the results of the comprehensive model.

The self-reports for lifestyle behaviour in the study pose a potential weakness, as the Lifestyle Risk indicators are then subject to potential measurement errors and self-recall biases. As wearables and fitness trackers become more commonplace in the future, health behaviour data from such devices would form more

objective sources for feeding into the model.

## Conclusion

We have presented a general model illustrating modifiable pathways from Lifestyle Risk to Diabetes Risk via the mediating factor of Physiological Load, and have tested it using large datasets from Wave 4 and Wave 5 of the Indonesian Family Life Survey. Non-modifiable sociodemographic covariates were included in the model for accounting of their effects, while focusing on what is amenable for health outcomes. The advantages of this model are that it groups and summarises a myriad of risk factors by behaviour components and a physiological index, and illustrates parsimonious and modifiable pathways that could be applied in public health efforts for diabetes (or chronic disease) prevention.

## Abbreviations

95% CI  
95% confidence intervals  
BIC  
Bayesian information criterion  
BMI  
Body mass index  
CFI  
Comparative fit index  
CRP  
C-reactive protein  
DALY  
Disability-Adjusted Life Years  
DBP  
Diastolic blood pressure  
ELISA  
Enzyme-linked immunosorbent assay  
HbA1c  
Glycosylated haemoglobin  
HDL  
High-density lipoprotein  
hsCRP  
high-sensitivity CRP

## Declarations

### Ethics approval and consent to participate

This study used publicly available, de-identified data, and was approved by the Singapore Health Services (SingHealth) Centralised Institutional Review Board (reference number: 2020/2021).

### Consent for publication

Not applicable

### Availability of data and materials

The data used in this study have been made available to the public on the RAND Social and Economic well-being website [<https://www.rand.org/well-being/social-and-behavioral-policy/data.html>]

### Competing interests

The authors declare that they have no competing interests.

### Funding

YLH and VSYL are funded by a grant, NMRC/CG/C027/2017.

### Author contributions

YLH and JT conceptualized the study and the overall study design. YLH, VSYL and MRH designed the statistical models. VSYL performed the data analysis, with major contributions from MRH and YLH. GJL gave significant inputs on the analysis and interpretation of the dietary intake. YLH and VSYL co-wrote the manuscript. All authors reviewed and agreed on the final version of the manuscript.

## Acknowledgements

Not applicable

## References

- Amalia B, Cadogan SL, Prabandari YS, Filippidis FT. Socio-demographic inequalities in cigarette smoking in Indonesia, 2007 to 2014. *Prev Med.* 2019;123:27–33. <https://doi.org/10.1016/j.ypmed.2019.02.025>.
- Ananta A, Arifin EN, Hasbullah MS, Handayani NB, Pramono W. *Demography of Indonesia's Ethnicity*. Singapore: Institute of Southeast Asian Studies; 2015.
- Aune D, ó Hartaigh B, Vatten LJ. Resting heart rate and the risk of type 2 diabetes: A systematic review and dose-response meta-analysis of cohort studies. *Nutrition Metabolism Cardiovascular Diseases.* 2015;25(6):526–34. <https://doi.org/10.1016/j.numecd.2015.02.008>.
- Aune D, Norat T, Leitzmann M, Tonstad S, Vatten LJ. Physical activity and the risk of type 2 diabetes: A systematic review and dose-response meta-analysis. *Eur J Epidemiol.* 2015;30(7):529–42. <https://doi.org/10.1007/s10654-015-0056-z>.
- Baker P, Friel S. Processed foods and the nutrition transition: evidence from Asia. *Obes Rev.* 2014;15(7):564–77. <https://doi.org/10.1111/obr.12174>.
- Bakker EA, Sui X, Brellenthin AG, Lee D. Physical activity and fitness for the prevention of hypertension. *Curr Opin Cardiol.* 2018;33(4):394–401. <https://doi.org/10.1097/HCO.0000000000000526>.
- Bardenheier BH, Bullard KMK, Caspersen CJ, Cheng YJ, Gregg EW, Geiss LS. Novel use of structural equation models to examine factors associated with prediabetes among adults aged 50 years and older: National health and nutrition examination survey 2001–2006. *Diabetes Care.* 2013;36(9):2655–62. <https://doi.org/10.2337/dc12-2608>.
- Barnes AS. (2011). The epidemic of obesity and diabetes: Trends and treatments. *Texas Heart Institute Journal, 38(2)*, 142–144. Retrieved from .
- Bays HE, Chapman RH, Grandy S. The relationship of body mass index to diabetes mellitus, hypertension and dyslipidaemia: comparison of data from two national surveys. *Int J Clin Pract.* 2007;61(5):737–47. <https://doi.org/10.1111/j.1742-1241.2007.01336.x>.
- Bei B, Seeman TE, Carroll JE, Wiley JF. (2017). Sleep and Physiological Dysregulation: A Closer Look at Sleep Intraindividual Variability. *Sleep, 40(9)*. <https://doi.org/10.1093/sleep/zsx109>.
- Bellou V, Belbasis L, Tzoulaki I, Evangelou E. Risk factors for type 2 diabetes mellitus: An exposure-wide umbrella review of meta-analyses. *PLoS ONE.* 2018;13(3):e0194127. <https://doi.org/10.1371/journal.pone.0194127>.
- Blasio A, Di, Donato F, Di, Mazzocco C. (2005). Guidelines for the data processing and analysis of the International Physical Activity Questionnaire. Retrieved May 18, 2020, from <http://www.ipaq.ki.se>.
- Bollen KA, Bauldry S. Three Cs in Measurement Models: Causal Indicators, Composite Indicators, and Covariates. *Psychol Methods.* 2011;16(3):265–84. <https://doi.org/10.1037/a0024448>.
- Buxton OM, Marcelli E. Short and long sleep are positively associated with obesity, diabetes, hypertension, and cardiovascular disease among adults in the United States. *Soc Sci Med.* 2010;71(5):1027–36. <https://doi.org/10.1016/j.socscimed.2010.05.041>.
- Byrne BM. Choosing Structural Equation Modeling Computer Software: Snapshots of LISREL, EQS, AMOS, and Mplus. In: R. Hoyle, *Handbook of Structural Equation Modeling*. New York: The Guilford Press; 2012. pp. 307–24.
- Camelo LV, Giatti L, Neves JAB, Lotufo PA, Benseñor IM, Chor D, ... Barreto SM. Life Course Socioeconomic Position and C-Reactive Protein: Mediating Role of Health-Risk Behaviors and Metabolic Alterations. The Brazilian Longitudinal Study of Adult Health (ELSA-Brasil). *PLoS ONE.* 2014;9(10):e108426. <https://doi.org/10.1371/journal.pone.0108426>.
- Carlsson AC, Andreasson N, A., & Wändell PE. Poor self-rated health is not associated with a high total allostatic load in type 2 diabetic patients - But high blood pressure is. *Diabetes Metabolism.* 2011;37(5):446–51. <https://doi.org/10.1016/j.diabet.2011.03.005>.
- Cattell RB. The Scree Test For The Number Of Factors. *Multivar Behav Res.* 1966;1(2):245–76. [https://doi.org/10.1207/s15327906mbr0102\\_10](https://doi.org/10.1207/s15327906mbr0102_10).
- Chobanian AV, Bakris GL, Black HR, Cushman WC, Green LA, Izzo JL, ... Roccella EJ. The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure: The JNC 7 Report. *J Am Med Assoc.* 2003;289(19):2560–72. <https://doi.org/10.1001/jama.289.19.2560>.
- Christiani Y, Byles JE, Tavener M, Dugdale P. Gender Inequalities in Noncommunicable Disease Risk Factors among Indonesian Urban Population. *Asia Pac J Public Health.* 2015;28(2):134–45. <https://doi.org/10.1177/1010539515626265>.
- Clark AJ, Dich N, Lange T, Jennum P, Hansen ÅM, Lund R, Rod NH. Impaired sleep and allostatic load: Cross-sectional results from the Danish Copenhagen Aging and Midlife Biobank. *Sleep Med.* 2014;15(12):1571–8. <https://doi.org/10.1016/j.sleep.2014.07.013>.

- De Rekeneire N, Peila R, Ding J, Colbert LH, Visser M, Shorr RI, ... Harris TB. Diabetes, hyperglycemia, and inflammation in older individuals: The Health, Aging and Body Composition study. *Diabetes Care*. 2006;29(8):1902–8. <https://doi.org/10.2337/dc05-2327>.
- Di Milia L, Vandelanotte C, Duncan MJ. The association between short sleep and obesity after controlling for demographic, lifestyle, work and health related factors. *Sleep Med*. 2013;14(4):319–23. <https://doi.org/10.1016/j.sleep.2012.12.007>.
- Dinno A. (2018). *paran: Horn's Test of Principal Components/Factors, version 1.5.2*. Retrieved from <https://cran.r-project.org/web/packages/paran/index.html>.
- Finkelstein EA, Chay J, Bajpai S. The economic burden of self-reported and undiagnosed cardiovascular diseases and diabetes on Indonesian households. *PLoS ONE*. 2014;9(6):99572. <https://doi.org/10.1371/journal.pone.0099572>.
- Fogelholm M. Physical activity, fitness and fatness: relations to mortality, morbidity and disease risk factors. A systematic review. *Obes Rev*. 2010;11(3):202–21. <https://doi.org/10.1111/j.1467-789X.2009.00653.x>.
- Fogelholm M, Kukkonen-Harjula K. Does physical activity prevent weight gain - a systematic review. *Obes Rev*. 2000;1(2):95–111. <https://doi.org/10.1046/j.1467-789x.2000.00016.x>.
- Forrester SN, Leoutsakos JM, Gallo JJ, Thorpe RJ, Seeman TE. Association between allostatic load and health behaviours: A latent class approach. *J Epidemiol Community Health*. 2019;73(4):340–5. <https://doi.org/10.1136/jech-2018-211289>.
- Franklin SB, Gibson DJ, Robertson PA, Pohlmann JT, Fralish JS. Parallel Analysis: a method for determining significant principal components. *J Veg Sci*. 1995;6(1):99–106. <https://doi.org/10.2307/3236261>.
- Gay JL, Salinas JJ, Buchner DM, Mirza S, Kohl HW, Fisher-Hoch SP, McCormick JB. Meeting Physical Activity Guidelines is Associated with Lower Allostatic Load and Inflammation in Mexican Americans. *J Immigr Minor Health*. 2015;17(2):574–81. <https://doi.org/10.1007/s10903-013-9950-1>.
- Gress TW, Nieto FJ, Shaha E, Wofford MR, Brancati FL. Hypertension and Antihypertensive Therapy as Risk Factors for Type 2 Diabetes Mellitus. *N Engl J Med*. 2000;342(13):905–12. <https://doi.org/10.1056/NEJM200003303421301>.
- Gunzler D, Chen T, Wu P, Zhang H. Introduction to mediation analysis with structural equation modeling. *Shanghai Archives of Psychiatry*. 2013;25(6):390–4. <https://doi.org/10.3969/j.issn.1002-0829.2013.06.009>.
- Hayashi T, Tsumura K, Suematsu C, Endo G, Fujii S, Okada K. High normal blood pressure, hypertension, and the risk of type 2 diabetes in Japanese men: The Osaka Health Survey. *Diabetes Care*. 1999;22(10):1683–7. <https://doi.org/10.2337/diacare.22.10.1683>.
- Heo M, Kim RS, Wylie-Rosett J, Allison DB, Heymsfield SB, Faith MS. Inverse Association between Fruit and Vegetable Intake and BMI even after Controlling for Demographic, Socioeconomic and Lifestyle Factors. *Obesity Facts*. 2011;4(6):449–55. <https://doi.org/10.1159/000335279>.
- Herningtyas EH, Hu P, Edenfield M, Strauss J, Crimmins E, Witoelar F, ... Sikoki B. Dried Blood Spot User's Guide for the Indonesia Family Life Survey, Wave. 5 (No. WR-1143/6-NIA/NICHD): RAND Corporation; 2017.
- Hoyle RH. *Handbook of Structural Equation Modeling*. New York: The Guilford Press; 2012.
- Hu P, Edenfield M, Potter A, Kale V, Risbud A, Williams S, ... Seeman T. Validation and modification of dried blood spot-based glycosylated hemoglobin assay for the longitudinal aging study in India. *American Journal of Human Biology*. 2015;27(4):579–81. <https://doi.org/10.1002/ajhb.22664>.
- Hussain MA, Al Mamun A, Reid C, Huxley RR. (2016). Prevalence, awareness, treatment and control of hypertension in Indonesian adults aged  $\geq 40$  years: Findings from the Indonesia Family Life Survey (IFLS). *PLoS ONE*, 11(8). <https://doi.org/10.1371/journal.pone.0160922>.
- International Diabetes Federation. (2015). *IDF Diabetes Atlas, 7th edition*. Brussels, Belgium: International Diabetes Federation.
- International Diabetes Federation. (2019). *IDF Diabetes Atlas, 9th edition*. Brussels, Belgium: International Diabetes Federation.
- Jarrin DC, Alvaro PK, Bouchard MA, Jarrin SD, Drake CL, Morin CM. Insomnia and hypertension: A systematic review. *Sleep Med Rev*. 2018;41:3–38. <https://doi.org/10.1016/j.smr.2018.02.003>.
- Juster RP, McEwen BS, Lupien SJ. Allostatic load biomarkers of chronic stress and impact on health and cognition. *Neurosci Biobehav Rev*. 2010;35(1):2–16. <https://doi.org/10.1016/j.neubiorev.2009.10.002>.
- Kahn SE, Hull RL, Utzschneider KM. Mechanisms linking obesity to insulin resistance and type 2 diabetes. *Nature*. 2006;444(7121):840–6. <https://doi.org/10.1038/nature05482>.
- Kaiser HF. The Application of Electronic Computers to Factor Analysis. *Educ Psychol Measur*. 1960;20(1):141–51. <https://doi.org/10.1177/001316446002000116>.
- Kaminsky LA, Montoye AHK. (2014). Physical activity and health: What is the best dose? *Journal of the American Heart Association*, 3(5). <https://doi.org/10.1161/JAHA.114.001430>.

- Kasapis C, Thompson PD. The effects of physical activity on serum C-reactive protein and inflammatory markers: A systematic review. *J Am Coll Cardiol*. 2005;45(10):1563–9. <https://doi.org/10.1016/j.jacc.2004.12.077>.
- Lidfeldt J, Nerbrand C, Samsioe G, Agardh CD. Women living alone have an increased risk to develop diabetes, which is explained mainly by lifestyle factors. *Diabetes Care*. 2005;28(10):2531–6. <https://doi.org/10.2337/diacare.28.10.2531>.
- Lin X, Xu Y, Pan X, Xu J, Ding Y, Sun X, ... Shan PF. Global, regional, and national burden and trend of diabetes in 195 countries and territories: an analysis from 1990 to 2025. *Sci Rep*. 2020;10(1):14790. <https://doi.org/10.1038/s41598-020-71908-9>.
- Mattei J, Noel SE, Tucker KL. A Meat, Processed Meat, and French Fries Dietary Pattern Is Associated with High Allostatic Load in Puerto Rican Older Adults. *J Am Diet Assoc*. 2011;111(10):1498–506. <https://doi.org/10.1016/j.jada.2011.07.006>.
- McEwen BS. Protective and Damaging Effects of Stress Mediators. *N Engl J Med*. 1998a;338(3):171–9. <https://doi.org/10.1056/NEJM199801153380307>.
- McEwen BS. Stress, Adaptation, and Disease: Allostasis and Allostatic Load. *Ann N Y Acad Sci*. 1998b;840(1):33–44. <https://doi.org/10.1111/j.1749-6632.1998.tb09546.x>.
- McEwen BS, Seeman T. Protective and damaging effects of mediators of stress. Elaborating and testing the concepts of allostasis and allostatic load. *Ann N Y Acad Sci*. 1999;896(1):30–47. <https://doi.org/10.1111/j.1749-6632.1999.tb08103.x>.
- McEwen BS, Stellar E. Stress and the Individual: Mechanisms Leading to Disease. *Arch Intern Med*. 1993;153(18):2093–101. <https://doi.org/10.1001/archinte.1993.00410180039004>.
- McEwen BS, Wingfield JC. The concept of allostasis in biology and biomedicine. *Horm Behav*. 2003;43(1):2–15. [https://doi.org/10.1016/S0018-506X\(02\)00024-7](https://doi.org/10.1016/S0018-506X(02)00024-7).
- Micha R, Shulkin ML, Peñalvo JL, Khatibzadeh S, Singh GM, Rao M, ... Mozaffarian D. (2017). Etiologic effects and optimal intakes of foods and nutrients for risk of cardiovascular diseases and diabetes: Systematic reviews and meta-analyses from the nutrition and chronic diseases expert group (NutriCoDE). *PLoS ONE*, 12(4). <https://doi.org/10.1371/journal.pone.0175149>.
- Miller RG. (1981). *Simultaneous Statistical Inference* (2nd ed.). <https://doi.org/10.1007/978-1-4613-8122-8>.
- Mills KT, Stefanescu A, He J. The global epidemiology of hypertension. *Nat Rev Nephrol*. 2020;16(4):223–37. <https://doi.org/10.1038/s41581-019-0244-2>.
- Monterrosa AE, Haffner SM, Stern MP, Hazuda HP. Sex difference in lifestyle factors predictive of diabetes in Mexican-Americans. *Diabetes Care*. 1995;18(4):448–56. <https://doi.org/10.2337/diacare.18.4.448>.
- Muthén LK, Muthén BO. *Mplus User's Guide*. 8th ed. Los Angeles: Muthén & Muthén; 2017.
- Ng R, Sutradhar R, Yao Z, Wodchis WP, Rosella LC. Smoking, drinking, diet and physical activity - Modifiable lifestyle risk factors and their associations with age to first chronic disease. *Int J Epidemiol*. 2020;49(1):113–30. <https://doi.org/10.1093/ije/dyz078>.
- Nishida C, Barba C, Cavalli-Sforza T, Cutter J, Deurenberg P, Darnton-Hill I, ... Zimmet P. Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. *The Lancet*. 2004;363(9403):157–63. [https://doi.org/10.1016/S0140-6736\(03\)15268-3](https://doi.org/10.1016/S0140-6736(03)15268-3).
- Paradis AM, Godin G, Pérusse L, Vohl MC. Associations between dietary patterns and obesity phenotypes. *International Journal of Obesity*. 2009;33(12):1419–26. <https://doi.org/10.1038/ijo.2009.179>.
- Pearl J. The Causal Foundations of Structural Equation Modeling. In: R. Hoyle, *Handbook of Structural Equation Modeling*. New York: The Guilford Press; 2012. pp. 68–91.
- Pengpid S, Peltzer K. The Prevalence of Underweight, Overweight/Obesity and Their Related Lifestyle Factors in Indonesia, 2014–15. *AIMS Public Health*. 2017;4(6):633–49. <https://doi.org/10.3934/publichealth.2017.6.633>.
- Persson P-G, Carlsson S, Svanstrom L, Ostenson C-G, Efendic S, Grill V. Cigarette smoking, oral moist snuff use and glucose intolerance. *J Intern Med*. 2000;248(2):103–10. <https://doi.org/10.1046/j.1365-2796.2000.00708.x>.
- Petrovic D, Pivin E, Ponte B, Dhayat N, Pruijm M, Ehret G, ... Stringhini S. Sociodemographic, behavioral and genetic determinants of allostatic load in a Swiss population-based study. *Psychoneuroendocrinology*. 2016;67:76–85. <https://doi.org/10.1016/j.psyneuen.2016.02.003>.
- Poulter N. (2003). Global risk of cardiovascular disease. *Heart*, 89(SUPPL. 2), ii2–ii5. [https://doi.org/10.1136/heart.89.suppl\\_2.ii2](https://doi.org/10.1136/heart.89.suppl_2.ii2).
- Quatromoni PA, Copenhafer DL, D'Agostino RB, Millen BE. Dietary patterns predict the development of overweight in women. *J Am Diet Assoc*. 2002;102(9):1239–46. [https://doi.org/10.1016/s0002-8223\(02\)90275-0](https://doi.org/10.1016/s0002-8223(02)90275-0).
- Reimers A, Knapp G, Reimers C-D. Effects of Exercise on the Resting Heart Rate: A Systematic Review and Meta-Analysis of Interventional Studies. *Journal of Clinical Medicine*. 2018;7(12):503. <https://doi.org/10.3390/jcm7120503>.

- Reutrakul S, Van Cauter E. Sleep influences on obesity, insulin resistance, and risk of type 2 diabetes. *Metabolism: Clinical Experimental*. 2018;84:56–66. <https://doi.org/10.1016/j.metabol.2018.02.010>.
- Revelle W. (2019). *psych: Procedures for Psychological, Psychometric, and Personality Research, Northwestern University, Evanston, Illinois, USA, version 1.9.12*. Retrieved from <https://cran.r-project.org/package=psych>.
- Ridker PM. (2003). C-reactive protein: a simple test to help predict risk of heart attack and stroke. *Circulation*, 108(12). <https://doi.org/10.1161/01.CIR.0000093381.57779.67>.
- Roman-Urrestarazu A, Ali FMH, Reka H, Renwick MJ, Roman GD, Mossialos E. Structural equation model for estimating risk factors in type 2 diabetes mellitus in a Middle Eastern setting: Evidence from the STEPS Qatar. *BMJ Open Diabetes Research Care*. 2016;4(1):e000231. <https://doi.org/10.1136/bmjdr-2016-000231>.
- RStudio Team. (2020). *RStudio: Integrated Development Environment for R*. Retrieved from <https://www.rstudio.com/>.
- Schulze MB, Hu FB. Primary prevention of diabetes: What can be done and how much can be prevented? *Annu Rev Public Health*. 2005;26:445–67. <https://doi.org/10.1146/annurev.publhealth.26.021304.144532>.
- Seccareccia F, Pannoza F, Dima F, Minoprio A, Menditto A, Lo Noce C, Giampaoli S. Heart Rate as a Predictor of Mortality: The MATISS Project. *Am J Public Health*. 2001;91(8):1258–63. <https://doi.org/10.2105/AJPH.91.8.1258>.
- Seeman T, Epel E, Gruenewald T, Karlamangla A, McEwen BS. Socio-economic differentials in peripheral biology: Cumulative allostatic load. *Ann N Y Acad Sci*. 2010;1186(1):223–39. <https://doi.org/10.1111/j.1749-6632.2009.05341.x>.
- Seixas AA, Vallon J, Barnes-Grant A, Butler M, Langford AT, Grandner MA, ... Jean-Louis G. Mediating effects of body mass index, physical activity, and emotional distress on the relationship between short sleep and cardiovascular disease. *Medicine*. 2018;97(37):e11939. <https://doi.org/10.1097/MD.00000000000011939>.
- Slopen N, Non A, Williams DR, Roberts AL, Albert MA. Childhood Adversity, Adult Neighborhood Context, and Cumulative Biological Risk for Chronic Diseases in Adulthood. *Psychosom Med*. 2014;76(7):481–9. <https://doi.org/10.1097/PSY.0000000000000081>.
- Sotos-Prieto M, Bhupathiraju SN, Falcón LM, Gao X, Tucker KL, Mattei J. A Healthy Lifestyle Score Is Associated with Cardiometabolic and Neuroendocrine Risk Factors among Puerto Rican Adults. *The Journal of Nutrition*. 2015;145(7):1531–40. <https://doi.org/10.3945/jn.114.206391>.
- StataCorp. *Stata Statistical Software: Release 14*. College Station: StataCorp LP; 2015.
- Statistics Indonesia - Badan Pusat Statistik - BPS, National Population and Family Planning Board - BKKBN/Indonesia, Kementerian Kesehatan - Kemenkes - Ministry of Health/Indonesia, and I. I. (2013). *Indonesia Demographic and Health Survey 2012*. Jakarta, Indonesia: BPS, BKKBN, Kemenkes, and ICF International.
- Stephoe A, Hackett RA, Lazzarino AI, Bostock S, La Marca R, Carvalho LA, Hamer M. Disruption of multisystem responses to stress in type 2 diabetes: Investigating the dynamics of allostatic load. *Proc Natl Acad Sci USA*. 2014;111(44):15693–8. <https://doi.org/10.1073/pnas.1410401111>.
- Strauss J, Witoelar F, Sikoki B. (2016). *The Fifth Wave of the Indonesia Family Life Survey: Overview and Field Report* (No. WR-1143/1-NIA/NICHD). <https://doi.org/10.7249/wr1143.1>.
- Suryadinata L, Arifin EN, Ananta A. *Indonesia's Population: Ethnicity and Religion in a Changing Political Landscape*. Singapore: Institute of Southeast Asian Studies; 2003.
- The National Institute of Health Research and Development. *National report on basic health research, RISKESDAS*. Indonesia Ministry of Health, Jakarta; 2013.
- Tripathy JP, Thakur JS, Jeet G, Chawla S, Jain S, Pal A, ... Saran R. Prevalence and risk factors of diabetes in a large community-based study in North India: results from a STEPS survey in Punjab, India. *Diabetology Metabolic Syndrome*. 2017;9(1):1–8. <https://doi.org/10.1186/s13098-017-0207-3>.
- Tripathy JP, Thakur JS, Jeet G, Jain S. Structural equation modeling to identify the risk factors of diabetes in the adult population of North India. *Tropical Medicine Health*. 2018;46(1):23. <https://doi.org/10.1186/s41182-018-0104-y>.
- United Nations Statistics Division. (2015). *Population by sex and urban/rural residence*. Retrieved January 13, 2021, from Demographic Statistics Database website: <http://data.un.org/Data.aspx?d=POP&f=tableCode%3A1>.
- van Draanen J, Prelip M, Upchurch DM. Consumption of fast food, sugar-sweetened beverages, artificially-sweetened beverages and allostatic load among young adults. *Preventive Medicine Reports*. 2018;10:212–7. <https://doi.org/10.1016/j.pmedr.2017.11.004>.
- Wang X, Bao W, Liu J, Ouyang YY, Wang D, Rong S, ... Liu LG. Inflammatory markers and risk of type 2 diabetes: A systematic review and meta-analysis. *Diabetes Care*. 2013;36(1):166–75. <https://doi.org/10.2337/dc12-0702>.

West SG, Taylor AB, Wu W. Model Fit and Model Selection in Structural Equation Modeling. In: R. Hoyle, Handbook of Structural Equation Modeling. New York: The Guilford Press; 2012. pp. 209–31.

Willi C, Bodenmann P, Ghali WA, Faris PD, Cornuz J. Active smoking and the risk of type 2 diabetes: A systematic review and meta-analysis. *J Am Med Assoc.* 2007;298(22):2654–64. <https://doi.org/10.1001/jama.298.22.2654>.

World Health Organization. (2011a). Global atlas on cardiovascular disease prevention and control. Retrieved September 9, 2020, from World Health Organization website: [http://www.who.int/cardiovascular\\_diseases/publications/atlas\\_cvd/en/](http://www.who.int/cardiovascular_diseases/publications/atlas_cvd/en/).

World Health Organization. (2011b). Global status report on noncommunicable diseases 2010. Retrieved August 31, 2020, from World Health Organization website: [http://www.who.int/nmh/publications/ncd\\_report2010/en/](http://www.who.int/nmh/publications/ncd_report2010/en/).

World Health Organization. (2011c). Use of Glycated Haemoglobin (HbA1c) in the Diagnosis of Diabetes Mellitus: Abbreviated Report of a WHO Consultation (WHO/NMH/CHP/CPM/11.1). Retrieved January 8, 2021, from World Health Organization website: [https://www.who.int/diabetes/publications/diagnosis\\_diabetes2011/en/](https://www.who.int/diabetes/publications/diagnosis_diabetes2011/en/).

World Health Organization. (2016). Global report on diabetes. Retrieved December 14, 2020, from World Health Organization website: <https://www.who.int/publications/i/item/9789241565257>.

World Health Organization. (2018). Projections of mortality and causes of death, 2016 to 2060. Retrieved March 19, 2021, from World Health Organization website: [https://www.who.int/healthinfo/global\\_burden\\_disease/projections/en/](https://www.who.int/healthinfo/global_burden_disease/projections/en/).

World Health Organization. (2020). WHO Guidelines on Physical Activity and Sedentary Behaviour. Retrieved January 8, 2021, from World Health Organization website: <https://www.who.int/publications/i/item/9789240015128>.

## Figures

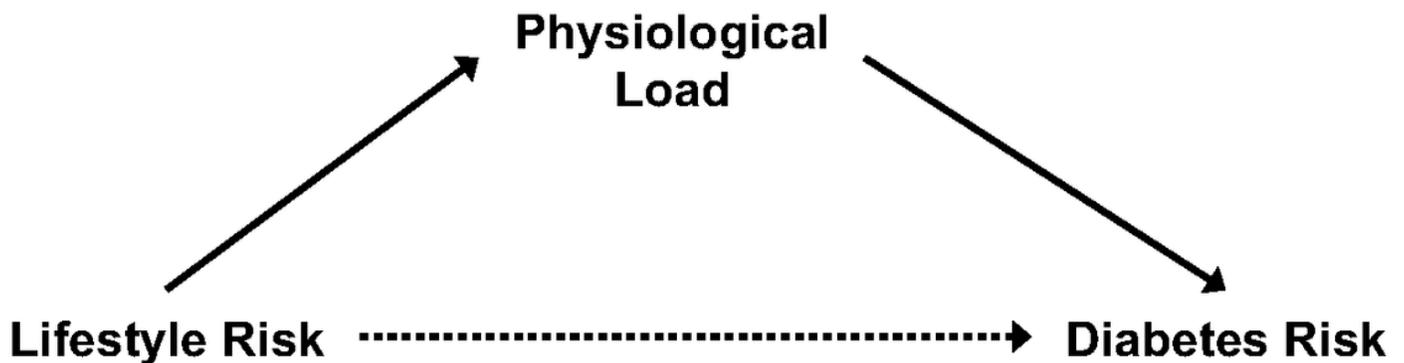
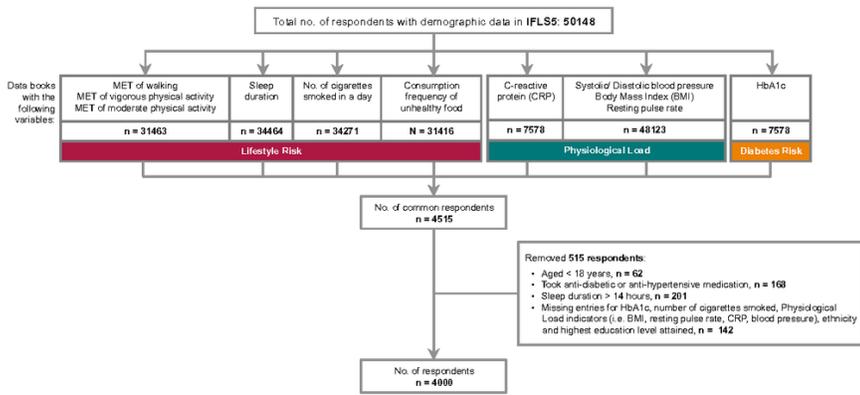


Figure 1

Illustration of the hypothesized pathways between Lifestyle Risk, Physiological Load and Diabetes Risk. Lifestyle Risk has a positive but indirect effect on Diabetes Risk through the mediator of Physiological Load; Lifestyle Risk has a direct and positive effect on the more downstream Physiological Load; Physiological Load has a direct and positive effect on the primary outcome of Diabetes Risk. Note: Straight line with one arrowhead denotes positive and direct effect; dashed line with one arrowhead denotes positive but indirect effect.

A



B

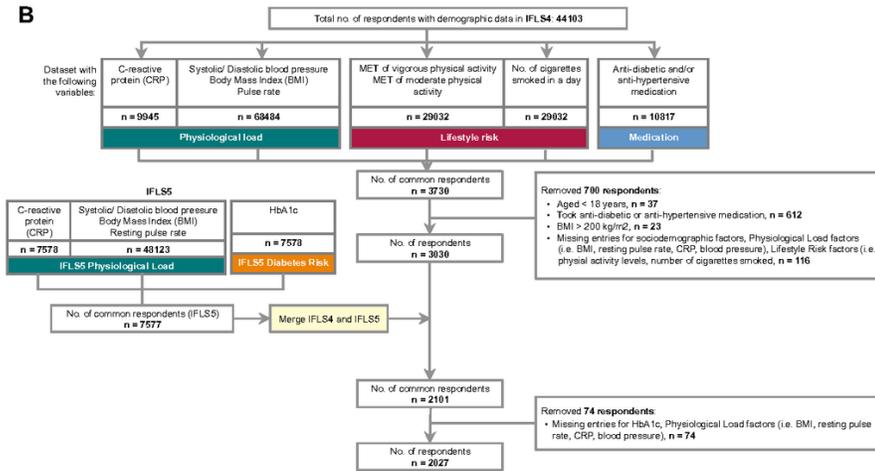
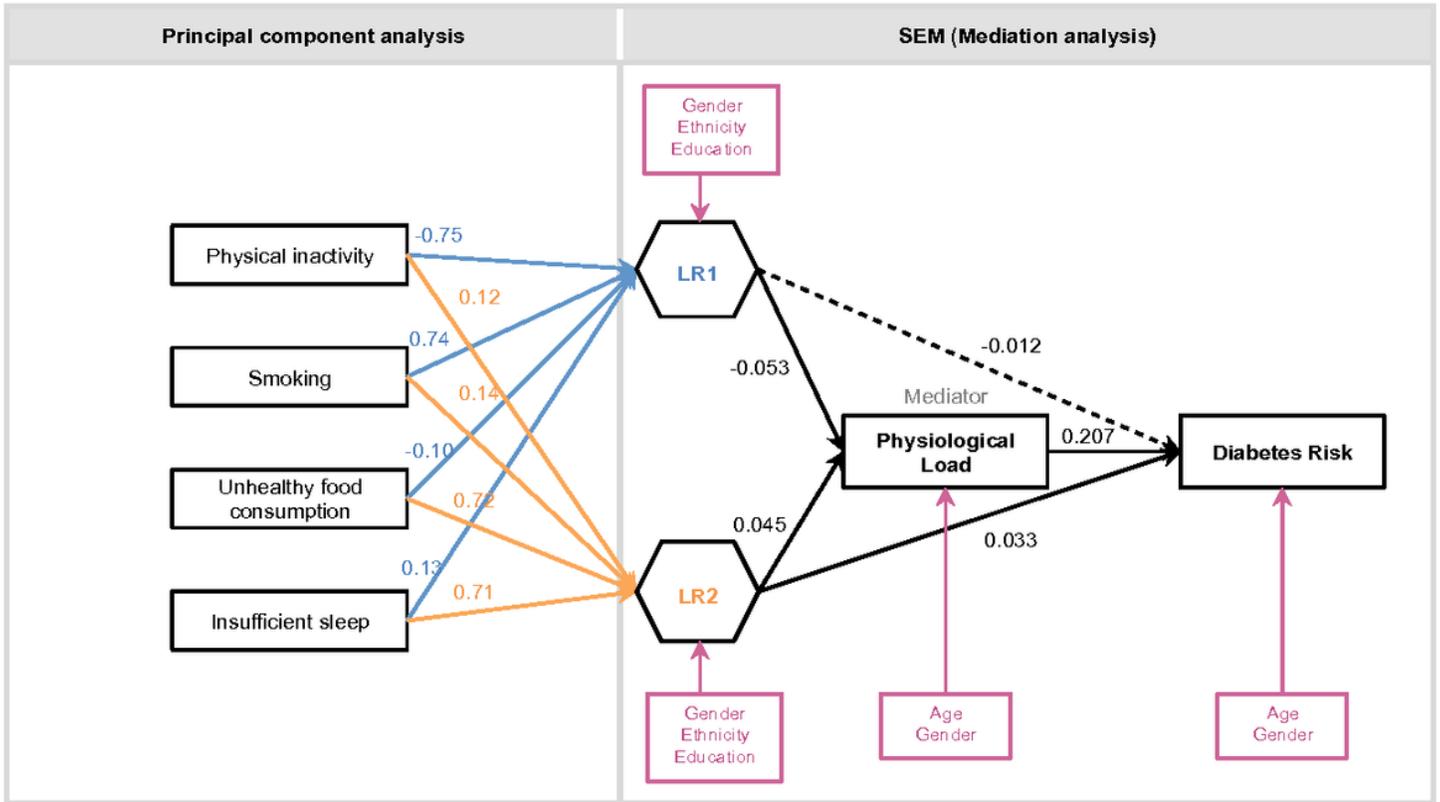


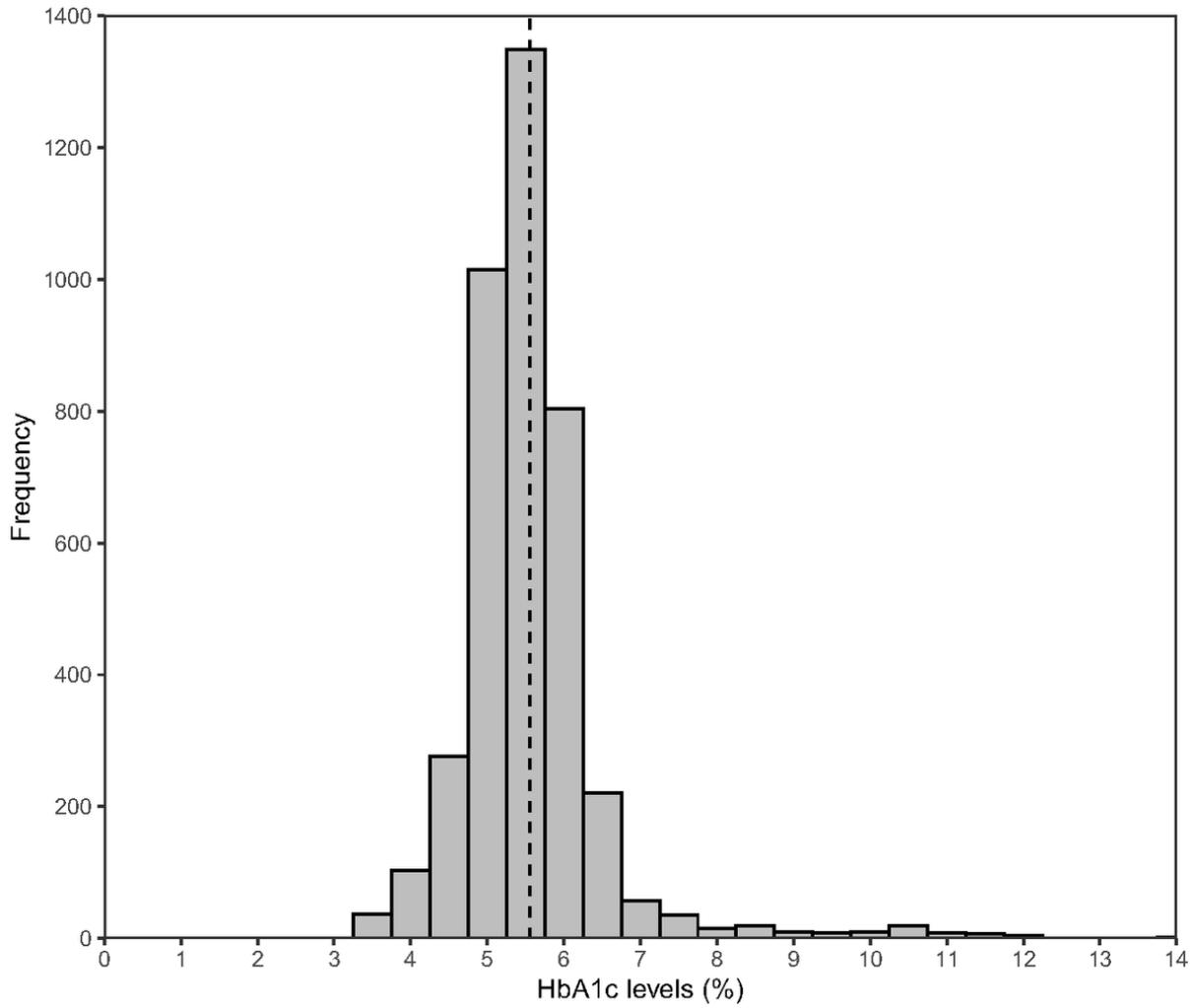
Figure 2

Flow chart of the study respondents in Model A and B. Figure 2A illustrates the flow chart of study respondents in Model A (IFLS5 data). The final study sample for Model A included 4000 respondents aged 18 and above; Figure 2B illustrates the flow chart of study respondents in Model B (IFLS4 and 5 data). The final study sample for Model B included 2027 respondents aged 18 and above. Note: In Figure 2A and 2B, the sum of items in exclusion criteria does not add up to 515 respondents and 700 respondents respectively as there were respondents who met multiple exclusion criteria.



**Figure 3**

The principal component analysis model and SEM mediation analysis in Model A. Lifestyle Risk composite variables, mediator Physiological Load and Diabetes Risk (HbA1c) outcome were from IFLS5 data. Socio-demographic variables, age, gender, ethnicity and highest education level were controlled for as covariates of Lifestyle Risk, Physiological Load and Diabetes Risk. Statistically significant socio-demographic covariates were presented in the figure. Note: Box indicates observed variable; hexagon indicates composite variables. Straight line with one arrowhead denotes significant effect; dashed line with one arrowhead denotes insignificant effect. Component loadings in principal component analysis were rounded off to 2 decimal places; Pathway loadings in SEM model were rounded off to 3 decimal places.



**Figure 4**  
Distribution of HbA1c levels (%) among 4000 respondents aged 18 and above. Note: Dashed line indicates the mean. Minimum HbA1c level = 3.50%; Maximum HbA1c level = 14.0%; Mean HbA1c level = 5.55%.

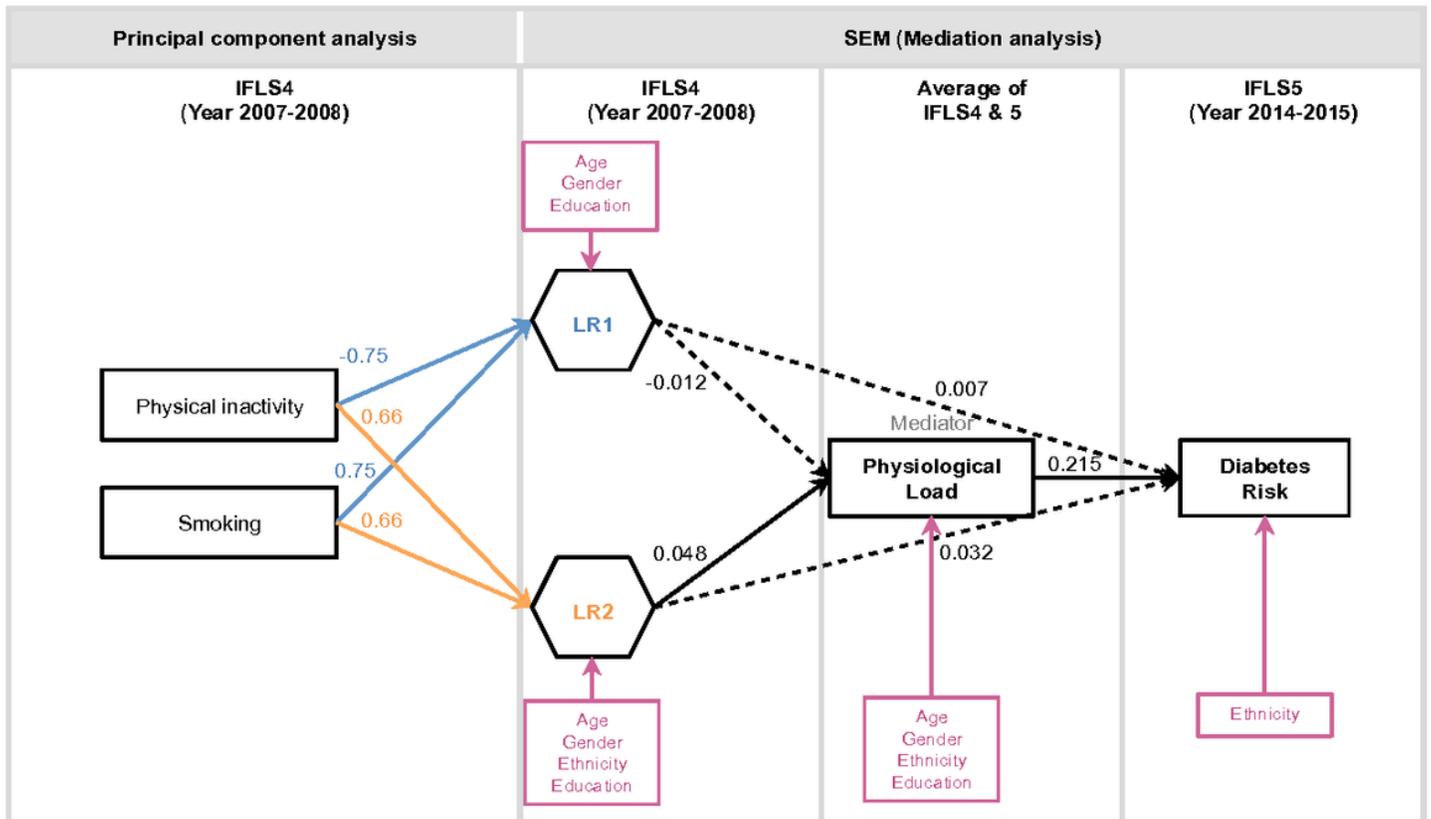


Figure 5

The principal component analysis model and SEM mediation analysis in Model B. Lifestyle Risk composite variables were from IFLS4 data, mediator Physiological Load from IFLS4 and IFLS5 data, and Diabetes Risk (HbA1c) outcome from IFLS5 data. Socio-demographic variables, age, gender, ethnicity and highest education level from IFLS4, were controlled for as covariates of Lifestyle Risk, Physiological Load and HbA1c outcome. Statistically significant socio-demographic covariates were presented in the figure. Note: Box indicates observed variable; hexagon indicates composite variables. Straight line with one arrowhead denotes significant effect; dashed line with one arrowhead denotes insignificant effect. Component loadings in principal component analysis were rounded off to 2 decimal places; Pathway loadings in SEM model were rounded off to 3 decimal places.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile1.pdf](#)
- [Additionalfile2.pdf](#)
- [Additionalfile3.pdf](#)
- [Additionalfile4.pdf](#)
- [Additionalfile5.pdf](#)
- [Additionalfile6.pdf](#)
- [Additionalfile7.pdf](#)