Analysis of obstructive sleep apnea, diabetes mellitus type 2, and prediabetes at dr. Cipto Mangunkusumo hospital using partial least squares

F E Puspitaningtyas¹, T Siswantining²* and T Kamelia³

¹,²Department of Mathematics, Indonesian University, Indonesia
³Division of Pulmonology, Department of Internal Medicine, Indonesian University, Indonesia

*Corresponding author: titin@sci.ui.ac.id

Abstract. This research aims to know the relationship model of OSA, DM type 2, and prediabetes simultaneously. Data used in this research is primary data which obtained by direct examination to patients with OSA at RSCM. The sampling method used in this research is non-probability sampling, with the number of samples is 205 patients. Partial Least Squares (PLS) method is used to obtain the purpose of this research. OSA, DM type 2, and prediabetes are dependent variables. Moreover, the independent variables are gender, age, blood pressure, obesity, and sleep hygiene. The risk of OSA is determined using the Berlin Questionnaire. DM type 2 and prediabetes are determined using fasting blood glucose (FBG). Using significance 0.1, we prove that OSA is directly affected by obesity and sleep hygiene. DM type 2 is directly affected by prediabetes, and indirectly affected by gender, age, obesity, and OSA. Gender, age, and OSA have direct effect to prediabetes, meanwhile, sleep hygiene has indirect effect to prediabetes. While obesity has direct and indirect effect to prediabetes, through OSA.

1. Introduction
Obstructive Sleep Apnea (OSA) is recognized as major cause of morbidity and mortality worldwide [1]. OSA is one of sleep-disordered breathing syndrome, indicated by upper airway collapse during sleep. This collapse can lead to oxygen desaturation and disrupted sleep. Symptoms of OSA are common, but patients are often unaware, because OSA occurs when patients are sleeping. Patients with OSA usually snore loudly and breathing pauses during sleep, have drowsiness during the day. OSA is a common disease that effects 2 – 4% adult population around the world [2]. Prevalence of OSA in Indonesia does not yet have a definite number, but a research on the prevalence of OSA in Indonesia have been done. Susanto et al in 2016, conducted OSA prevalence based on Berlin questionnaire on traffic police in East Jakarta, and found 17,2% of 93 respondents had OSA possibilities. Useful and simple tool for detecting OSA is Berlin questionnaire. Berlin questionnaire has 3 sections: score of the snore; daytime sleepiness during daily activities; medical history, demographic, and anthropometric measures [4].

Obesity is one of the major risk factors for OSA. Some measurements such as Body Mass Index (BMI), neck circumference (NC), and waist circumference (WC) can be used to measure obesity [5]. Another risk factors for OSA are gender, age, blood pressure (BP), and sleep hygiene (SH). Sleep hygiene can be reflected by smoking habit, use of light, having meal 3 hours before bedtime, and coffee consumption.
OSA can be linked to another disease, such as Diabetes Mellitus (DM). A study found that 30.1% of OSA patients had DM type 2 [6]. OSA through intermittent hypoxemia, a condition in which the human body temporarily loses sufficient oxygen supply to the blood, can cause several effects such as oxidative stress, systemic inflammation, activation of HPA axis, which caused insulin resistance that leads to Diabetes Mellitus (DM) type 2 [7]. Further more, OSA and DM type 2 have a common risk factor, such as obesity ([5],[8]). Based on the relationship and their risk factors, the hypothesis is described in Figure 1 below. Building the relationship model between OSA, DM type 2, prediabetes, and their risk factors simultaneously is the novelty of this paper.

Figure 1. Model Hypothesis between OSA, DM Type 2, and Prediabetes and Their Risk Factors

2. Research Variables
Variables in this research consist of: status of DM type 2 and prediabetes; OSA (categorized as high risk and low risk based on the Berlin questionnaire); gender; age; blood pressure; obesity and sleep hygiene as latent variables.

3. Experimental Method
Population in this research are patients with OSA at RSCM who met the inclusion criteria and did not include exclusion criteria. This research has been reviewed and approved by the Health Research Ethics Committee, Faculty of Medicine, Universitas Indonesia – Cipto Mangunkusumo Hospital. The sample of this research consisted of patients with OSA in RSCM. The number of samples in this research were 205 patients, who were taken using purposive sampling. The method for data analysis Partial Least Squares (PLS), to determine the relationship between these three diseases simultaneously, so the purposes of this research can be achieved.

3.1. Partial Least Squares (PLS) Method
Partial Least Squares (PLS) is a used to explain a relation between multiple blocks of variables [9]. PLS also can be used to see the pattern of relationship between latent variables (variables that can’t be measured directly, but can be measured by measuring their indicators) and measurable variables. In PLS, there are two relationship models, namely inner model (between latent variables) and outer model (latent variables and its indicators). In PLS, based on indicators that explain the latent variables there are two types of constructs used, namely reflective and formative. The direction of the relationship in the reflective construct is from the latent variable to the indicator variable. Meanwhile, in formative constructs, the direction is from indicator variables to the latent variable [10].

The outer model equation for the reflective construct is [10]:

\[ E(X_{jp} | LV_j) = \lambda_{0jp} + \lambda_{jp}LV_j \]  

(1)
where: $LV_j$ is latent variable – $j$; $X_{jp}$ is indicator $p$ of latent variable $j$; $\lambda_{jp}$ is value of $X_{jp}$ if the value $LV_j$ equals to zero; $\lambda_{jp}$ is loading indicator $p$ of latent variable $j$ if the construct is reflective.

The outer model equation for the formative construct if latent variable $j$ has $p$ indicators is [10]:

$$E(LV_j | X_{j1}, X_{j2}, ..., X_{jp}) = v_{0j} + \sum_{p=1}^{p} v_{jp} X_{jp}$$

(2)

where: $LV_j$ is latent variable – $j$; $X_{jp}$ is indicator $p$ of latent variable $j$; $v_{0j}$ is value of $X_{jp}$ if the value $LV_j$ equals to zero for every $p$; $v_{jp}$ is weight indicator $p$ of latent variable $j$ if the construct is formative.

PLS does not require strong assumptions, such as distribution assumption, sample size, and measurement scale. Variables involved in PLS can have nominal, ordinal, interval, and ratio scales [11]. In PLS models, between latent variables and between latent variables and its indicator variables are described using a path diagram.

3.2. Parameters in PLS

Parameters that will be estimated in this research are outer loading and path coefficient. Suppose there are $j$ latent variables, with the value of the latent variable denoted by $y_j$, $j = 1, 2, ..., J$. Each latent variable $j$ is measured by indicator variables, with the $p$-th indicator variable notation on the latent variable $j$ is $x_{j1}, x_{j2}, ..., x_{jp}$. The number of indicator variables ($p$) for each latent variable can vary.

Suppose that the number of iterations needed to converge is $r$ with $r = 1, 2, ..., R$. The iteration procedure in parameter estimation in the PLS model is carried out with the following stages [10]:

- Give any weight. The iteration process starts by giving any weight to each $p$-th indicator on the latent variable $j$, called $w_{jp}$, with $p = 1, 2, ..., P$. Iteration starts with $w_{jp} = 1$.
- Standardization the value of indicator variables. The $p$-th indicator on the latent variable $j$ ($X_{jp}$) needs to be standardized so that the resulting value has one variance. The measurement scale used in PLS can vary, therefore the standardization the value of indicator variables is used to equalize the units of the measurement scale of the indicator variables $X_{jp}$ is the value of $X_{jp}$ that has been standardized.
- Calculate external approximation, that is estimation latent variable $j$ as a weighted linear combination of the standardized $X_{jp}$ with $w_{jp}$ weights.
- Standardization the value of latent variable $y_j$. Just like $X_{jp}$, latent variable $y_j$ will also be standardized.
- Determine inner weight. In this step, inner weight will be calculated, which is $e_{ij}$ for each relationship if the latent variable $j$ is influenced by latent variable $i$.
- Calculate internal approximation. Internal approximation of the latent variable denoted by $z_j$ is a linear weighted combination of the latent variable $i$ with the standardized external approximation.
- Standardization value of latent variable $z_j$. Call the standardized latent variable $z_j$ is $\tilde{z}_j$. $\tilde{z}_j$ is an $n \times 1$ sized vector containing the internal approximation value of the standardized latent variable $j$, which is $\tilde{z}_j$.
- Calculate the new outer weight again. The new outer weight for reflective construct can be calculated by the following formula:

$$w_j = (\tilde{z}'_j \tilde{z}_j)^{-1} \tilde{z}'_j \tilde{X}_j$$

(3)

- Checking the convergence of $w_{jp}$. For each $r$-th iteration with $r = 1, 2, ..., R$ the convergence of $w_{jp}$ in $r$-th iteration and $w_{jp}$ in $(r - 1)$-th iteration is being checked. Convergence checking is done
with the terms $|w_{jp}^{(r)} - w_{jp}^{(r-1)}| < 10^{-5}$ [10]. If the condition has been fulfilled for each $j$ and $p$, then the iteration is stopped. However, if these conditions have not been fulfilled, then step 3 – 8 will be repeated until the convergence of $w_{jp}$ is met.

- Calculate the estimated value of the latent variable. After the final weight that has converged is obtained, it will calculate the estimated value of the latent variable with the following formula:

$$LV_j = y_j = \sum_{p=1}^{P} w_{jp} \hat{x}_{jp}$$

(4)

- Calculate estimated path coefficients. Suppose $\hat{\beta}_{jl}$ is an estimate for $\beta_{jl}$. $\hat{\beta}_{jl}$ can be calculated with the following formula:

$$\hat{\beta}_{jl} = (\hat{y}'_i \hat{y}_i)^{-1} \hat{y}'_i \hat{y}_i$$

(5)

- Calculate the estimated outer loading. Outer loading is calculated as the correlation between the latent variable and its indicators. Outer loading can be calculated by the formula:

$$\hat{\lambda}_{jp} = \text{cor}(\hat{X}_{jp}, \hat{y}_j)$$

(6)

3.3. Partial Least Squares (PLS) Method

Model evaluation in PLS consists of evaluating the outer model and inner model.

3.3.1 Outer Model Evaluation

- **Convergent Validity.** Convergent validity corresponds to the principle that the indicator variable of a latent variable should be highly correlated. Test convergent validity in PLS for reflective indicators rated based on outer loading, ideally is 0.7, but 0.6 is still acceptable [11].

- **Discriminant Validity.** Discriminant validity relates to the principle that indicators from different latent variables are not highly correlated. Discriminant validity test in PLS is assessed based on cross loading, namely the correlation between indicator variables from a latent variable with another latent variable. Therefore, discriminant validity can be fulfilled if outer loading $>\text{cross loading}$, with cross loading $\geq 0.7$.

- **Composite Reliability.** In addition to validity testing, PLS also conducts reliability testing. Reliability testing is used to measure the consistency of indicators in making measurements. Reliability testing in PLS uses composite reliability. Composite reliability can be calculated by formula [11]:

$$\rho_j = \frac{(\sum_{p=1}^{P} \lambda_{jp})^2}{(\sum_{p=1}^{P} \lambda_{jp})^2 + \sum_{p=1}^{P} (1 - \lambda_{jp}^2)}$$

where $\rho_j$ is composite reliability for latent variable $j$ with $p$ indicators. Indicator variables from reflective constructs are said to be reliable when composite reliability $\geq 0.7$.

3.3.2 Inner Model Evaluation

- **Path Coefficient.** Path coefficient is a parameter that states the magnitude of the effect of an independent variable on the dependent variable. Path coefficient is obtained from the iterations described earlier. Testing the significance of path coefficient can be done using the bootstrapping method. If testing for path coefficient using the bootstrap method is significance, then the inner model can be said to be fit.

- $R^2$ or the coefficient of determination indicated that the variability of the dependent variable can be explained by the independent variable. $R^2$ is used to measure the level of variation in changes
independent variable to the dependent variable. According to Sanchez (2013), $R^2$ is classified as follows: low if $R^2 < 0.3$; moderate for $0.3 < R^2 < 0.6$; and high for $R^2 \geq 0.6$.

4. Result and Discussion

The data obtained in this research were analyzed using Partial Least Squares method with SmartPLS Professional (Trial Version). The dependent variables in this research were OSA, DM type 2, and prediabetes, while independent variables are gender, age, obesity, sleep hygiene, and blood pressure.

![Figure 2. The Best Model of the Relationship of OSA, DM type 2, and Prediabetes Simultaneously](image)

After evaluation of the outer and inner model, the best model is obtained for the model of the relationship between OSA, DM type 2, and prediabetes. Based on Figure 2 above, we can conclude that OSA is directly affected by obesity and sleep hygiene. DM is affected directly by prediabetes and indirectly affected by gender, obesity, age, and OSA. Prediabetes is influenced directly by gender, age, and OSA. Prediabetes is indirectly affected by sleep hygiene. Obesity can affect prediabetes in direct and indirect ways through OSA. The indicator that most strongly reflects obesity is OB 1, namely BMI with outer loading 0.957. Meanwhile, the use of light (SH 2) is an indicator reflects sleep hygiene the most, with outer loading 0.730. The results obtained are almost in accordance with the hypotheses that have been made previously. But there are also several factors that do not affect and are not related, such as blood pressure with OSA, DM, and prediabetes.

Using SmartPLS Professional (Trial Version) we also get the coefficient of determination ($R^2$), shown in Table 1. Table 1 shows that 94.7% variation of DM, according to Figure 3, can be well explained by prediabetes, while 11.6% variation of OSA, according to Figure 3, can be explained by obesity and sleep hygiene, and 14% variation of prediabetes, according to Figure 3, can be explained by gender, obesity, OSA, and age.

| Variable | $R^2$ |
|----------|-------|
| DM       | 0.947 |
| OSA      | 0.116 |
PREDM 0.140
5. Conclusion
Based on the analysis using PLS method, we get the best model of the relationship of OSA, DM type 2, and prediabetes simultaneously in Figure 3. In Figure 3, we can conclude that OSA affects Prediabetes directly, by that means OSA can also affect DM type 2 indirectly through Prediabetes. Risk factors for OSA are obesity and sleep hygiene, meanwhile, risk factors for prediabetes are gender, age, obesity, and OSA. We can see from Figure 3, prediabetes affects DM type 2 directly, it means risk factors that affect prediabetes also affect DM type 2 indirectly. OSA, DM type 2, and prediabetes shared the same risk factors, such as obesity and sleep hygiene.

Acknowledgements
This research is supported by PITTA UI 2018 Research Granted.

References
[1] Pataka A, Daskalopoulou E, Kalamaras G, Fekete Passa K, and Argyropoulou P 2014 *Sleep Medicine* 15(7) 776
[2] Jordan A S, McSharry D G and Malhotra A 2014 *The Lancet* 383(9918) 736
[3] Susanto A 2016 *J Respir Indo* 36
[4] Khaledi-Paveh B, Khazaie H, Nasouri M, Ghadami M R and Tahmasian M 2016 *Basic Clin. Neuroscience* 7 (1) 43
[5] Pinto J 2011 *Brazilian J of Otorhinolaryngology* 77 516
[6] Mok Y, Tan C W, Wong H S, How C H, Tan K L A and Hsu P P 2017 *Singapore Med J.* 58(4) 179
[7] Reutrakul S and Mokhlesi B 2017 *CHEST* 152(5) 1070
[8] Awasthi A, Rao C R, Hedge D S and K Rao N 2017 *Prev Med HYG* 58(1) E56
[9] Attia S, Versloot C J, Voskuijl W, Vliet S J, Di Giovanni V, Zhang L, Richardson D, Bourdon C, Netea M G, Berkley J A, van Rheenen P F and Bandsma R H 2016 *Am. J. Clin. Nutr.* 104(5) 1441
[10] Sanchez G 2013 *PLS path modeling with R* (Berkelev: Trowchez Editions)
[11] Ehrmann T, Josef W, Gerard C and George H 2013 *Network governance alliances, cooperatives, and franchise chains* (Physica-Verlag: Heidelberg)