使用整合之多基因風險分數預測糖尿病患者罹患憂鬱症

Integrated Polygenic Risk Scores to Predict Depressive Disorder Among Patients with Type 2 Diabetes

Internship Institution :

Department of Medical Research, Taichung Veterans General Hospital

Student: Yun-Chen Chung **Advisors:** Tzu-Hung Hsiao, Ph.D, Ching-Heng Lin, Ph.D, Prof. Tzu-Pin Lu

Abstract

It has been proved that patients with type 2 diabetes mellitus (T2DM) have a higher risk of developing depressive disorder (DD) [1]. People who suffer from both diseases are faced with negative impacts such as greater mortality. Polygenic Risk Score (PRS) is an innovative tool to predict the disease of individuals by calculating weighted score of specific single nucleotide polymorphisms (SNPs). PRS can serve as a reminder to help people prevent DD and get early diagnosis. Thus, to reform single PRS, this study aims to construct a PRSmix model that combine multiple PRS into a model to predict the onset of DD among T2DM.

Material and Methods

Study Design

We used genetic data in Taiwan Precision Medicine Initiative (TPMI) and medical records of International Classification of Diseases (ICDcode) in Taichung Veterans General Hospital from 2004 to 2022. Cases are T2DM patients with DD, except those who took the medicine that might cause weight gain, hypertension, and hyperglycemia two years before they were diagnosed with T2DM. As it takes time to develop DD, controls were T2DM patients who were diagnosed with T2DM before the last month of medical records and did not have DD. We matched the data by sex and age and set the ratio of 1:10, keeping 3,707 samples for genome-wide association study (GWAS), and 4,774 samples for PRS and PRSmix analysis.

Statistical Analysis

We implemented nearest neighbor matching using logistic regression propensity score. To ensure the comparability of cases and controls, we carried out T-test and Chi-squared test. Afterwards, we conducted a genome-wide association study (GWAS) using PLINK and LocusZoom. PRS were analyzed as a continuous or categorical variable (using 90% and 95% quantile cut point respectively). We used T-test to test the differences of PRS among cases and control groups, and logistic regression model to estimate the effect size of each PGSs. Afterwards, step-wise logistic regression were used to enhance the performance of the PRSmix. McFadden's Pseudo R-squared, receiver operating characteristic (ROC) curve and area under the ROC curve (AUC) were utilized to evaluated the performance of models.

Result

Study Population

Female comprised 67% of samples, with average age 63. The quality of matching was good since the p-value were large (Table 1), and the distribution of propensity scores and matched variables were similar in cases and controls. (Figure not shown)

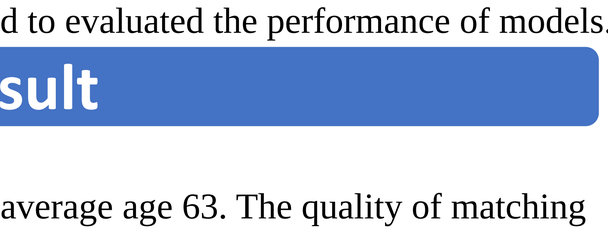
Table 1. The characteristic of study population

	total	case	control	P-value
GWAS	n = 3707	n = 337	n = 3370	
Female sex (%)	2508 (67.66)	228 (67.66)	2280 (67.66)	1
Mean (SD) age, years	63.6018 (12.2683)	63.5608 (12.3074)	63.6059 (12.2662)	0.9489
PRS Analysis	n = 4774	n = 434	n = 4340	
Female sex (%)	3201 (67.05)	291 (67.05)	2910 (67.05)	1
Mead (SD) age, years	63.7193 (12.4189)	63.6797 (12.4487)	63.7233 (12.4173)	0.9446

Genome-Wide Association Analysis

We did not find any significant SNPs in DD patients among T2MD population (Figure 1). Failing to recognize the SNPs associated with DD, we then used the existing PRS proposed by other researchers.

Fig 1. Manhattan Plot of GWAS of 337 cases with both T2DM and DD, and 3370 controls with T2DM only.



PRS Analysis

We filtered out 105 PGS that were used to predict diabetes or depression by other researchers. We carried out T-tests to find 7 PGS, most of which were used to predict T2DM, had significant mean differences between cases and controls (table 2). We randomly split the data into training and testing data by the ratio 7:3. Using logistic regression, we evaluated the effect size of PRS adjusted for sex and age. McFadden's Pseudo R-squared ranging between 0.2-0.4 indicates a good model, but all of our models scored lower than that. The AUC of the PGS (categorical and continuous) that were significant in logistic regression model were between 0.47 - 0.54, which was not very ideal. (table 3 and 4, continuous PRS not shown)

Table 2. Significant PGS in T-test

PGS	P-value
PGS000027	0.0432
PGS000330	0.0188
PGS000867	0.0437
PGS000846	0.0361
PGS000729	0.0090
PGS000852	0.0355
PGS000144	0.0384

Table 3. Significant PGS in logistic regression model, using 90% quantile cut point *R^2: McFadden's Pseudo R-squared

PGS	P-value	R^2	AUC (95% CI)
PGS000851	0.0365	0.0023	0.4792 (0.4310 - 0.4792)
PGS002426	0.0485	0.0025	0.5143 (0.4670 - 0.5143)
PGS000138	0.0092	0.0043	0.5122 (0.4635 - 0.5122)
PGS002659	0.0040	0.0041	0.5068 (0.4562 - 0.5068)

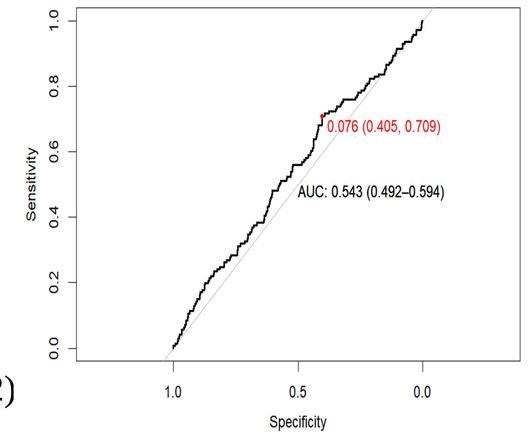
Table 4. Significant PGS in logistic regression model, using 95% quantile cut point *R^2: McFadden's Pseudo R-squared

PGS	P-value	R^2	AUC (95% CI)
PGS000851	0.0004	0.0023	0.4894 (0.4393 - 0.4894)
PGS000138	0.0099	0.0057	0.5198 (0.4717 - 0.5198)
PGS000032	0.0349	0.0039	0.5008 (0.4510 - 0.5009)
PGS000852	0.0041	0.0051	0.4913 (0.4426 - 0.4913)

PRSmix

We integrated the 7 PGS that were significant in T-tests and 4-5 PGS that were significant in logistic regression model. The 90% quantile groups worked the best, but the AUC is still extremely low (AUC=0.515, 95% CI: 0.471 - 0.559). We then tried a model of 7 PGS that were significant in T-tests only. The performance was improved, and it worked best when the PRS score variable were continuous (AUC=0.543, 95%CI: 0.492 - 0.594). (Figure 2)

Figure 2. ROC curve of step-wise logistic regression model, using 7 significant PGS in T-test as continuous variables



Discussion and Conclusion

This study developed several PRSmix models that combine multiple PRS together to enhance the prediction for DD among T2DM patients. There were some limitations. First, the sample size is too small to represent the general population. In addition, we did not consider other factors including lifestyles, environment [2], socioeconomic status, and comorbidity with other diseases, which play influential roles in developing DD. In conclusion, though the performance of our PRSmix were not good, further study can use the PGS we filtered out and integrate them with other variables irrelevant to genes or heredity to enhance the prediction.

Reference

1. Huang, C. J., Hsieh, H. M., Tu, H. P., Jiang, H. J., Wang, P. W., & Lin, C. H. (2018). Major depressive disorder in patients with type 2 diabetes mellitus: Prevalence and clinical characteristics. *Journal of affective disorders*, *227*, 141–148. https://doi.org/10.1016/j.jad.2017.09.044
2. Kendler, K. S., Karkowski, L. M., & Prescott, C. A. (1999). Causal relationship between stressful life events and the onset of major depression. *The American journal of psychiatry*, *156*(6), 837–841. https://doi.org/10.1176/ajp.156.6.837