World Academy of Science, Engineering and Technology
International Journal of Bioengineering and Life Sciences
Vol:19, No:09, 2025

Application of Principal Component Analysis and Ordered Logit Model in Diabetic Kidney Disease Progression in People with Type 2 Diabetes

Authors : Mequanent Wale Mekonen, Edoardo Otranto, Angela Alibrandi

Abstract : Diabetic kidney disease is one of the main microvascular complications caused by diabetes. Several clinical and biochemical variables are reported to be associated with diabetic kidney disease in people with type 2 diabetes. However, their interrelations could distort the effect estimation of these variables for the disease's progression. The objective of the study is to determine how the biochemical and clinical variables in people with type 2 diabetes are interrelated with each other and their effects on the progression of kidney diseases using advanced statistical methods. As a first step, principal component analysis was used to examine how the biochemical and clinical variables intercorrelate with each other, which helped us reduce a set of correlated biochemical variables to a smaller number of uncorrelated variables. Different types of ordered logit regression models were applied to determine the effect of principal components and clinical variables on the order-level response variable (progression of kidney function). These models were fitted within the framework of category comparison of the response variable (cumulative, stage, and adjacent) and under the assumption of proportionality. This retrospective cross-sectional study retrieved data from a type 2 diabetic cohort in a polyclinic hospital at the University of Messina, Italy. The principal component analysis yielded three uncorrelated components. These are principal component 1, which had negative loading of glycosylated hemoglobin, glycemia, and creatinine; principal component 2, which had negative loading of total cholesterol and low-density lipoprotein; and principal component 3, which had negative loading of high-density lipoprotein and positive loading of triglycerides. The cumulative odds model indicated that the first principal component (a linear combination of glycosylated hemoglobin, glycemia, and creatinine) was a strong and significant risk factor for the progression of kidney disease, with an effect or odds ratio of 0.423 (95% confidence interval, 0.32–0.52, P value = 0.000). However, because the variable principal component 1 does not meet the proportionality assumption, this effect is an invalid estimate. Alternative ordered logit models that allow the effect to vary across the progression level of kidney disease, such as the partial adjacent category models, the partial continuation ratio model, and the partial cumulative odds model, were used to obtain accurate effect estimates. The partial cumulative odds model, partial continuation ratio, and partial adjacent category suggested that clinical variables age, sex, body mass index, medication (metformin), and biochemical variables glycosylated hemoglobin, glycemia, and creatinine significantly impact the progression of kidney disease.

Keywords : diabetic kidney disease, ordered logit model, principal component analysis, type 2 diabetes
Conference Title : ICBBBB 2025 : International Conference on Biomathematics, Biostatistics, Bioinformatics and Bioengineering
Conference Location : Naples, Italy
Conference Dates : September 20-21, 2025