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 kidney disease progression through advanced statistical methods. First, principal component analysis was used to explore 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. Then, ordered logit regression models (cumulative, stage, and adjacent) were employed to assess the effect of biochemical and clinical variables on the order-level response variable (progression of kidney function) by considering the proportionality assumption for more robust effect estimation. 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, with negative loading of glycosylated haemoglobin, glycemia, and creatinine; principal component 2, with negative loading of total cholesterol and low-density lipoprotein; and principal component 3, with negative loading of high-density lipoprotein and a positive load of triglycerides. The ordered logit models (cumulative, stage, and adjacent) showed that the first component (glycosylated haemoglobin, glycemia, and creatinine) had a significant effect on the progression of kidney disease. For instance, the cumulative odds model indicated that the first principal component (linear combination of glycosylated haemoglobin, glycemia, and creatinine) had a strong and significant effect on the progression of kidney disease, with an effect or odds ratio of 0.423 (P value = 0.000). However, this effect was inconsistent across levels of kidney disease because the first principal component did not meet the proportionality assumption. To address the proportionality problem and provide robust effect estimates, alternative ordered logit models, such as the partial cumulative odds model, the partial adjacent category model, and the partial continuation ratio model, were used. These models suggested that clinical variables such as age, sex, body mass index, medication (metformin), and biochemical variables such as glycosylated haemoglobin, glycemia, and creatinine have a significant effect on 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