**Publication News 79 – 17 July 2023**

**Identification of risk factors for DPN using Big Data and Machine Learning**

***Aims:*** To determine the risk factors associated with DPN and chronic pain in DPN without *a priori* assumption in the UK Biobank (UKB) Study.
***Methods:*** This data analytics cohort study assessed 11,047 patients with diabetes recruited to the UKB study. This study included people with diabetes prospectively developing DPN two to ten years after baseline assessment at the UKB. All participants had a raft of clinical measurements and assays. DPN diagnosis was based on electronic data record coding on follow-up and painful-DPN on the basis of a score of 3 or greater on questions 1-7 on the DN4 questionnaire *at baseline*. Risk factors were determined with a machine learning algorithm to determine individual disease risk, and risk factor importance ranked.
***Results:*** DPN was diagnosed in 794 cases (7.19%) during 10-year follow up (incidence 12% in T1DM, mean age 57.6 years, 45.7% female and 6% in T2DM, 60.6 years, 69.5% male) and 40% classified as painful-DPN. Lower socioeconomic status, obesity, cystatin C, HbA1c, and CRP predicted DPN risk with differences in the rank of risk factors between T1DM and T2DM. On further examination of inflammatory variables, there were sex specific differences in immune cell involvement, such as neutrophil and monocyte count. In addition, CRP was a risk factor for painful-DPN. Finally, protective factors for DPN included IGF-1, Vitamin D and raised haemoglobin. The machine learning model showed discriminative performances with AUC >0.64 predictive capability.
***Conclusions:*** Lifestyle factors and blood biomarkers were identified as risk factors for DPN. Machine learning tools might have clinical utility to identify those at risk for developing DPN for targeted intervention.

**Comments.** This study highlights the benefits of 'big data', and indeed some of its flaws, through its use of the UKB. This study highlights known risk factors of DPN across T1DM and T2DM such as HbA1c and nephropathy, with slight differences between the two types of diabetes (Neutrophil-Lymphocyte Ratio and reduced IGF-1 predicting DPN in T2DM only). The model identified CRP as a strong risk factor for DPN and painful-DPN, strengthening the argument that systemic inflammation is important in the pathogenesis of DPN and painful-DPN. Moreover, the study identified sex-specific differences in risk factors, a novel observation that the pathophysiology of DPN may vary between men and women. There are some limitations of the study, largely due to the nature of the study and impracticability of more detailed testing/assessments. The case definition of DPN was based on medical coding, which is unreliable and insensitive, as only clinically evident cases would be diagnosed. Moreover, neuropathic pain was diagnosed at baseline, rather than follow-up, limiting the interpretation of the risk factors for painful-DPN. Further research is necessary to determine whether risk models, such as these, using big data have clinical utility to identify those at risk. Perhaps what is required first, is proving that there are effective treatments in T2DM for those at risk of DPN to prevent the disease.

**Gordon Sloan**

**Reference**. Allwright M, Karrasch JF, O'Brien JA, Guennewig B, Austin PJ. Machine learning analysis of the UK Biobank reveals prognostic and diagnostic immune biomarkers for polyneuropathy and neuropathic pain in diabetes. Diabetes Res Clin Pract. 2023 Jul;201:110725. doi: 10.1016/j.diabres.2023.110725. Epub 2023 May 19. PMID: 37211253.
https://www.diabetesresearchclinicalpractice.com/article/S0168-8227(23)00488-6/fulltext