**Aim:** this model aims to show my understanding of key principles within multiple binary outcome modelling and programming skill by developing and validating a multivariate bayesian probit prediction model.

**Data source**：the dataset is from the in-person practical week three of the Msc course, statistical modelling and inference for health. The dataset contains information on 200,000 observations between Jan 2009 and Dec 2018, and includes the following variables:

Sex: the sex of the individual

Smoking status: indicates whether the individual was or is a smoker (1=previous/ current smoker, 0=non-smoker)

Diabetes: indicates whether the individual has diabetes (1=diabetic, 0=not diabetic)

CKD: indicates whether the individual has chronic kidney disease (1=chronic kidney disease).

It is also available as “[DevelopmentData(1).csv](https://github.com/yuqiwa/Multivariate-Bayesian-probit-CPM/blob/main/DevelopmentData(1).csv)” on Github.

**Binary outcomes**:(1) Diabetes; (2) CKD.

**Predictors:** (1)Sex; (2)Smoking status.

**Missing data:** no missing data.

**Statistical methods:** we fitted this model using Bayesian inference, and parameter estimation were obtained naturally using MCMC methods.

**Evaluation:**report predictive performance in a random hold-out sample (30%).

|  |  |
| --- | --- |
|  | Joint risk performance |
| Calibration | Calibration intercept and slope |
| Discrimination | AUC |

**Result on model performance:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | Outcome | Calibration-in-the-large | Calibration slope | AUC |
| MPM | P11 | -0.03121294 | 1.119791 | 0.6350053 |
| MPM | P10 | 0.03940292 | -0.1419054 | 0.5043842 |
| MPM | P01 | 0.01609859 | 1.002756 | 0.6234049 |

Reference

Martin G P, Sperrin M, Snell K I E, et al. Clinical prediction models to predict the risk of multiple binary outcomes: a comparison of approaches[J]. Statistics in medicine, 2021, 40(2): 498-517.