Biomedical Data Science
2025, VOL. 5, NO. 3, 183-186
DOI: 10.47297/wspbdsWSP2752-630538.20250503

Analysis of Association Between Diabetes, Stroke and Hypertension Using Joint Model, Mixed Effects Model and Cox Model

Hanyan Yuan[1], Yi Zhang[2]

[1]School of International Relations, University of International Business and Economics, Beijing 100029

[2]School of International Development and Cooperation, University of International Business and Economics, Beijing 100029

ABSTRACT

Background: Diabetes, stroke and hypertension are all common diseases, and they are known to be associated with each other. People with hypertension seem to be at a higher risk at diabetes and stroke. Previous studies have used Cox proportional hazard models to analyze the association between repeated measures and each disease independently. However, previous studies fail to take into account on repeated measurements of blood pressure and time-to-event data for two diseases into one model to analyze their association. Also, separate analyses would cause bias and invalid inference on parameter estimates, since those previous analyses fail to capture the latent relationship among clinical outcomes. Methods: We used the joint model to evaluate the association between diabetes, stroke and hypertension on 5,000 UK Biobank participants. In joint model, we could link the longitudinal and time-to-event outcome through the random effects. We used linear mixed effect model and Cox proportional hazards model to do the model diagnostics for survival part and longitudinal part separately. Results: In the longitudinal sub-model, age at visit, body mass index (BMI), sex, ethnicity were all statistically significant ($p<0.05$). In the competing risks sub-model, BMI, sex and ethnicity were significantly associated with diabetes ($p<0.0001$), but not with stroke. Our results indicated that higher BMI and increased age were associated with an increased risk at diabetes, as well as a higher blood pressure. Male received higher risk at diabetes than female ($p<0.0001$). Non-white population received the higher risk than white. Discussion and Conclusion: Higher blood pressure contributed to higher risk at diabetes, but we didn't find significant evidence that higher blood pressure had an impact on stroke. One possible reason is that the participants had much lower occurrence rate at stroke. In the longitudinal data diagnostics, we found that the assumption of normal distribution of random effect may not be appropriate. In the survival data diagnostics, the Cox-Snell plot showed that Cox PH model is a good fit for modeling stroke, but not for diabetes. Therefore, further research on model diagnostics of joint model is necessary to take into account the latent association.

KEYWORDS

Joint modeling; Longitudinal data analysis; Linear mixed model; Cox proportional hazard model

1 Introduction

1.1 Background

Diabetes is a chronic disease in which your blood glucose levels are higher than usual. Glucose acts as the primary energy source for the cells that make up muscles and tissues of humans. Without glucose, there is no function in the body. There are two main types of diabetes: type I and type II. Type I diabetes can appear at any age, though it is most common among children and adolescents. While most type II diabetes develops after the age of 30, and half of the newly diagnosed patients with type II diabetes are over 40 years of age at the time of onset [1]. Type II diabetes accounts for 90 to 95 percent of total diabetes cases in the United States, while type I diabetes covers the remaining 5 to 10 percent [2]. Potential dangers of type II diabetes include: heart and vessel disease, nerve damage, kidney disease, eye disease and so on [1]. Consequently, the prevention of type II diabetes is critically vital to human health. In this study, we focused on type II diabetes. In 2015, a meta-analysis conducted by Connor A. Emdin looked at 4.1 million adults, who were free of diabetes and cardiovascular disease. This study concluded that people with high blood pressure were at a higher risk of developing type II diabetes [3]. A 20 mm Hg increase in systolic blood pressure (SBP) was associated with a 58% increased risk of new-onset diabetes in their study result. The occurrence of stroke is mainly due to cerebral vascular obstruction or rupture, stopping brain tissue from getting oxygen and nutrients [4]. The symptoms of a stroke include: can't say words clearly and can't understand others; feeling paralysis of the arms, legs or face; headache and lose balance to walk. Approximately 4.5 million people are dying from stroke each year in the world, with over 9 million stroke survivors. Nearly a quarter of men and nearly a fifth of women aged 45 are at risk of having a stroke if they live to the age of 85 [5]. Stroke is an emergency disease, so early detection and control are crucial. In 2004, Carlene M. M. Laweswas demonstrated that the expected epidemiological benefits of lowering blood pressure on stroke risk were broadly consistent across a range of population subgroups [6]. In their research, a reduction of 10 mm Hg in systolic blood pressure was associated with a reduced risk of stroke by about one-third, based on the data from randomized controlled trials at an average age of about 70 years. Also, evidence showed that greater benefit with a larger blood pressure reduction.

1.2 UK biobank data

UK Biobank is the world's largest and most influential data platform. Basing the introduction of the UK Biobank

website, the purpose of the UK government building this platform is to provide researchers with the world's largest and most comprehensive research resources of human information on genetic and environmental factors, to explore the links between genes, lifestyle and health, and to improve the prevention, diagnosis and treatment of a range of serious and life-threatening diseases – such as cancer, heart disease and stroke. It has made significant contributions to the advancement of modern medicine and treatment and has allowed numbers of scientific discoveries that improve human health. The UK Biobank study has collected 500,000 volunteers who were recruited between 2006 and 2010. These volunteers living in the UK were aged between 40 and 69 when they joined UK Biobank and would be followed at least 30 years thereafter. In order to understand how the individuals experienced diseases, the UK Biobank collected some information, such as lifestyles and the blood, urine and saliva samples, regularly with the consent of the volunteers to link up with their health records [7]. A random subset of the full data was taken in this analysis.

2 Methods

2.1 Data summary

The random subset of data included 5,000 volunteers. Two datasets were used in this analysis: repeated measurements of longitudinal data for each individual and time-to-event data. Longitudinal data recorded each volunteer's ID, gender (female 2721, male 2279), ethnicity (non-white 234, white 4766), date of birth, repeated measurements of blood pressure at each hospital visit for each person and their age at the time of the visit (mean 54.6). The frequency of blood pressure tests varied for each person, and each patient has a mean of seven visits (median 6, minimum 1, maximum 104).

Time-to-event data included the same cohort as people in the longitudinal data. Survival data mainly includes the following variables in addition to those mentioned in the longitudinal data: Individual's survival time for the competing risk (this time refers to the first event occurred); Competing risk indicator: 0, 1 and 2 represent right-censoring (N= 4554), diabetes (N=322), and stroke (N=124), respectively; ethnicity indicator: 0 and 1 represent white and non-white, respectively; Age at a first stroke: Age of the patient at the time of the first stroke occur; Stroke: 0 and 1 indicate patients with stroke (N=137) and without stroke (N=4863), respectively; Age at first diabetes: Age of the patient at the time of first diabetes occur; Diabetes: 0 and 1 indicates patients with diabetes (N=331) and without diabetes (N=4669), respectively; and BMI for each person. Table 2 shows the mean and standard deviation of survival time for diabetes only, stroke only and competing risk disease.

2.2 Shared-random effect joint model formulation and likelihood

The shared-random effect joint model consists of two sub-models: longitudinal component and survival component. Note that the longitudinal data and time-to-event data are independent conditional on all the covariates and random effects. After derivation, the likelihood for this joint model can be written as:

$$
\begin{aligned}
L(\Psi;\ Y, C) &\propto \prod_{i=1}^{n} f(Y_i, C_i|\psi) \\
&= \prod_{i=1}^{n} \int_b f(Y_i|C_i, b, \psi) f(C_i|b, \psi) f(b|\psi) db \\
&= \prod_{i=1}^{n} \int_b f(Y_i|b, \psi) f(C_i|b, \psi) f(b|\psi) db \\
&= \prod_{i=1}^{n} \int_b [\prod_{j=1}^{n_i} \frac{1}{\sqrt{2\pi\delta^2}} exp\{-\frac{1}{2\delta^2}(Y_{ij} - X_i(t_{ij})^T\beta - Z_i(t_{ij})^T b)^2\}] \\
&\quad \{\prod_{k=1}^{g} \lambda_k(T_i; X_i^{(2)}(T_i), b, \gamma_k, \nu_k)^{I(D_i=k)}\} \\
&\quad \times exp[-\int_0^{T_i} \{\sum_{k=1}^{g} \lambda_k(t; X_i^{(2)}(t), b, \gamma_k, \nu_k)\} dt] \times \frac{1}{\sqrt{2\pi|\Sigma|}} exp(-\frac{1}{2} b^T \Sigma^{-1} b) db
\end{aligned}
$$

Due to it is a joint model, which is hard to calculate the analytical solution, we can use the EM algorithm to estimate its solution. Here we focused on the model described above.

3 Results

3.1 Result of the joint model

We first fitted the data on the joint model, the results were shown as below (Table 1).

We set age at each visit as fixed and random effect covariate. In the longitudinal part, the p-values of all the fixed effects were smaller than 0.05, so they were all significantly associated with blood pressure. Also, the coefficients could be interpreted as the same as the linear mixed effect model, such as a unit increase in BMI is associated with a 0.71 unit increase in the mean of blood pressure, controlling for other variables. In the survival sub-model, all these fixed effects were significantly associated with diabetes but not significant to stroke. Plus, the coefficients here could be interpreted as, for example, a unit increase in BMI is associated with a 16% increase in the hazard ratio of diabetes. Looking at the association parameter's part, the first number of the association parameter indicates the event, note that 1 is diabetes

Table 1 Joint model results

Longitudinal	**coef**	**SE**	**95%Lower**	**95%Upper**	**p-values**
Fixed Effects					
Intercept	90.46	1.40	87.71	93.20	0.00
Age at each visit	0.37	0.02	0.33	0.41	0.00
BMI	0.71	0.03	0.65	0.78	0.00
Male	3.64	0.34	2.97	4.32	0.00
Non-white	-2.10	0.80	-3.67	-0.52	0.01
Survival	**exp(coef)**	**SE(coef)**	**95%Lower**	**95%Upper**	**p-values**
Fixed Effects					
BMI_diabetes	1.16	0.01	0.13	0.16	0.00
Male_diabetes	1.81	0.12	0.37	0.82	0.00
Non-white_diabetes	3.21	0.19	0.80	1.53	0.00
BMI_stroke	1.02	0.02	-0.02	0.06	0.32
Male_stroke	1.29	0.18	-0.11	0.62	0.17
Non-white_stroke	1.35	0.41	-0.50	1.11	0.46
Association Parameters	**Est.**	**SE(coef)**	**95%Lower**	**95%Upper**	**p-values**
vee_11	1.02	0.01	0.01	0.04	0.00
vee_12	2.87	0.43	0.22	1.89	0.01
vee_21	1.02	0.01	0.00	0.04	0.10
vee_22	2.55	0.70	-0.43	2.31	0.18

and 2 is stroke. The second number refers to the random intercept and random slope, the random slope here is the age of each visit. From these parameters, both the random intercept and random slope of diabetes were significant. The intercept of the association parameter means that the higher the blood pressure, the more risk to develop diabetes. The slope of the association parameter meaning that if the blood pressure goes up over time, the risk of diabetes will be increased. For stroke, the parameters were also positive, indicating that they were in the right direction, but they were not significant [8].

3.2 Model diagnostics

3.2.1 Linear mixed effects model

We started with a profile plot to show the trajectory of systolic blood pressure among 50 male and 50 female patients that were randomly selected from the dataset. There ' re not much differences in SBP between two groups. And then, we used a scatterplot to present the relationship between four basic ranges of BMI (underweight 16-18.5, healthy 18.5-25, overweight 25-30 and obese 30-40) and SBP. As can be seen, most people in the data were overweight.

From table 2, the results in the linear mixed effects model were almost the same as the longitudinal sub-model from the joint model.

Table 2 Linear mixed effects model results.

Fixed effects:

	Est.	**Std. Error**	**t-value**	**P-value**
Intercept	90.59	1.27	71.12	0.00
Age at each visit	0.37	0.02	22.68	0.00
BMI	0.72	0.03	20.82	0.00
Male	3.71	0.34	11.04	0.00
Non-white	-2.13	0.83	-2.57	0.01

All fixed variables were significant. Figure 1 showed the diagnostic plots for the linear fixed effects model. Figure 1a was a residual plot, which could be used to check if the residuals have a constant variance. The variance seemed to tend larger when the x-axis getting larger. Figure 1b was the Q-Q plot, checking the normality of the residual. The line suggests a non-linear trend. Figure 1c was the observed responses versus the fitted values. If the dots look elliptical then it is meaning that the model is fitted. In Figure 1c, the dots go apart when x getting larger. Figure 1d was the normal quantile plots for each random effect. Under the assumptions of a linear mixed effect model, each random effect term should be normally distributed. However, the random effects violated the assumptions.

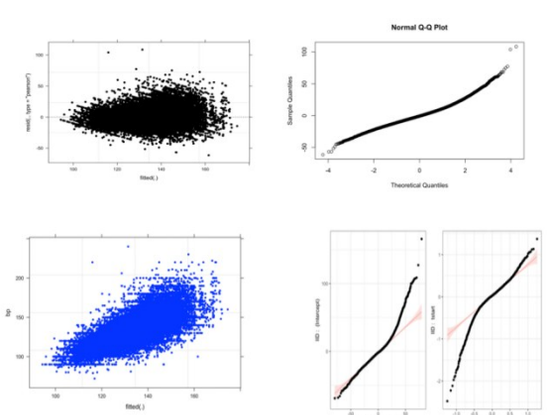

Figure 1 Model diagnostic plots for linear mixed effects model: 1a: Residual plot,1b: QQ plot, 1c: The observed responses versus the fitted values, 1d: Normal quantile plots for random effects

3.2.2 Cox proportional hazard model

We considered Cox PH model for type II diabetes and stroke, respectively. Firstly, sex, BMI and ethnicity, which are the fixed effects same as the linear mixed effects model, are considered as the covariates for Cox model. From the results, they looked similar to the joint model. All the parameters were significant. Then we did the model diagnostic plot for this cox model, which is a plot of residuals versus the estimated cumulative hazard of the residuals for the Cox model. Since the plot did not quite follow the 45-degree straight line, the Cox PH model did not seem to fit the data well.

Next, we fitted a Cox model to consider stroke only, the results were a little different from the joint model. The coefficient estimation part was similar, but the variable sex was significant in the Cox model, and it was not significant in the joint model. Other variables in the Cox PH model were not significant. The Cox-Snell residual plot for this model fitted the data better than the diabetes Cox PH model.

4 Discussion

We use separate analyses to do model diagnostics that Fast JM package cannot do. However, one limitation is that we cannot draw any conclusions about the latent association between two diseases and blood pressure. Stroke was not significantly associated with blood pressure, because of the lower event occurrence of stroke compared to that of diabetes in this cohort. If we are able to assess the complete data, it might help us to draw stronger statistical inference on parameter estimates. Therefore, further research on model diagnostics of joint model is necessary to take into account the latent association.

5 Conclusion

By conducting the joint model to the 5, 000 random participants from UK biobank, we concluded diabetes was positively significantly associated with blood pressure. But we did not have significant evidence to determine the relationship between blood pressure and stroke. Besides, BMI, sex, age at each visit and ethnicity might be reliable predictors for blood pressure; and BMI, sex and ethnicity were significant predictors for diabetes. In the longitudinal data diagnostics, we found that the assumption that random effect is normally distributed may not be appropriate. In the survival data diagnostics, the Cox-Snell plot shows that Cox PH model is a good fit for modeling stroke, but not for diabetes.

About the Author

1. Hanyan Yuan, Female, Han, Huaibei City, Anhui Province, Master Degree, Assistant Economist. Research Direction: Biostatistics, Educational Management.

2. Yi Zhang, Female, Han, Beijing, Master Degree, Junior Editor. Research Direction:Educational Management, Statistics.

References

[1] Mayo Foundation for Medical Education and Research. (2020, October 30). Diabetes. Mayo Clinic. https://www. mayoclinic. org/diseases-conditions/diabetes/symptoms-causes/syc-20371444.

[2] What Is Type 1 Diabetes? EndocrineWeb. (n.d.. https://www.endocrineweb.com/conditions/type-1-diabetes/type-1-diabetes#:~:text=Type%201%20diabetes%20makes%20up, immune%20response%20that%20triggers%20it.

[3] Emdin C. A., Anderson S. G., Woodward M., & Rahimi K. (2015, October 6). Usual Blood Pressure and Risk of New-Onset Diabetes: Evidence From 4.1 Million Adults and a Meta-Analysis of Prospective Studies. Journal of the American College of Cardiology. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4595710/.

[4] Mayo Foundation for Medical Education and Research. (2021, February 9). Stroke. Mayo Clinic. https://www. mayoclinic. org/diseases-conditions/stroke/symptoms-causes/syc-20350113#:~:text=A%20stroke%20occurs%20when%20the,brain%20damage%20and%20other%20complications.

[5] Charles D A Wolfe, The impact of stroke, British Medical Bulletin, Volume 56, Issue 2, 2000, Pages 275–286.

[6] Lawes C. M. M., Carlene M.M. Lawes From the Clinical Trials Research Unit, Bennett, D. A., Derrick A. Bennett From the Clinical Trials Research Unit, Feigin, V. L., Valery L. Feigin From the Clinical Trials Research Unit, Rodgers, A., Anthony Rodgers From the Clinical Trials Research Unit, & Lawes, C. to C. (2004, February 19). Blood Pressure and Stroke. Stroke.

[7] About us. (n.d.). https://www.ukbiobank.ac.uk/learn-more-about-uk-biobank/about-us.

[8] Robert M Elashoff, Gang Li, and Ning Li. A joint model for longitudinal measurements and survival data in the presence of multiple failure types. Biometrics, 64(3): 762-771, 2008.