國立臺灣大學公共衛生學院公共衛生學系

學士論文

Department of Public Health

College of Public Health

National Taiwan University

Bachelor's Thesis

以雙向孟德爾隨機化探究血小板與高血壓之因果關係

Elucidation of causal direction between Platelet count and Hypertension: a bi-directional Mendelian Randomization study

邱柏鈞

P0-Chun Chiu

指導教授: 盧子彬 博士

Advisor: Tzu-Pin Lu Ph.D.

中華民國 110 年 4 月

April, 2021

國立臺灣大學學士學位論文

口試委員會審定書

以雙向孟德爾隨機化探究血小板與高血壓之因果關係

Elucidation of causal direction between Platelet count and Hypertension: a bi-directional Mendelian Randomization study

本論文係邱柏鈞君（B06801012）在國立臺灣大學公共衛生學系完成之學士學位論文，於民國 110 年 4 月 20 日承下列考試委員審查通過及口試及格，特此證明

口試委員：_____

（指導教授）

_____ _____

_____ _____

系主任：_____

摘要

研究背景:

高血壓是許多重大慢性病的共同危險因子，也是目前世界衛生組織 (world health organization,WHO) 公布全球疾病負擔排名的首位。過去的研究顯示高血壓和血小板的數量有顯著相關，然而這些研究存在著樣本數偏少且難以進行隨機分派實驗去釐清彼此的因果關係，因此，本研究透過基因位點來剖析兩者之因果關係。

方法:

本研究資料來自台灣人體生物資料庫，包含 16,000 位年齡位於 30 歲到 70 歲的參與者。基因檢測使用的晶片為 Affymetrix Axiom TWB 1.0 晶片，共包含有 646,735 個單核苷酸多型性位點數據，我們透過文獻篩選特定基因位點，並使用孟德爾隨機化分析高血壓與血小板數量的因果關係。

結果:

以納入文獻選取出的 5 個基因為點作為工具變項，執行孟德爾隨機化後得到血小板數量對於高血壓具有正向且顯著的相關性 (odds ratio: 1.149, 95% CI: [-0.164,0.849], P=0.185)。

結論:

以台灣人體生物資料庫為研究資料並符合孟德爾隨機化的假設下，血小版數量對於高血壓有顯著因果關係，而血小版數量與高血壓間不存在雙向因果關係，可做為臨床上診斷高血壓的相關資訊。

Abstract

Background:

Observational associations between platelet activation and risk factors for hypertension are well established, but the exact nature of causality between them remains unclear.

Methods:

Clinical and genotype (single nucleotide polymorphisms (SNPs)) data from 15,996 healthy Taiwanese individuals aged between 30 and 70 years from the Taiwan Biobank project were included. We performed a bi-directional Mendelian randomization analysis using inverse variance weighting to estimate the causality of platelet count in hypertension. We used 65 platelet count-related SNPs and 6 hypertension-related SNPs as instrumental variables. Furthermore, to test for pleotropic effect of the instruments, sensitivity analyses was performed using the MR-Egger and weighted median methods.

Results:

This study provided evidence in support of a positive causal effect of platelet count on the

risk of hypertension ($odds \quad ratio: \ 1.149, \ 95\% CI: \ [1.131, \ 1.578], \ P < 0.05$), using the weighted median method. Significant causality of platelet count on hypertension was observed using the IVW method. However, no pleiotropy was observed for the instruments in the analyses.

Conclusions:

In this Taiwanese population with Han-Chinese ancestry, a significant positive causal relationship of platelet count on hypertension was revealed, whereas the causal effect of hypertension on platelet count was found to be non-significant. Platelet count could be used as a marker for the diagnosis of hypertension

Keywords: Mendelian randomization, bi-directional causal estimation, hypertension, platelet count, Taiwan Biobank

Contents

List of Figures

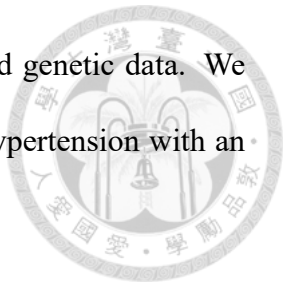
List of Tables

Chapter 1 INTRODUCTION

Hypertension is an important risk factor for major chronic diseases, such as cardiovascular disease, stroke, diabetes, and kidney disease. According to the Global Burden of Disease study, leading detailed risk factors for attributable disability-adjusted life-years (DALYs) were related to blood pressure. Hypertension is a multi-factorial disease[1] [2] [3], and some of the previous studies confirmed that there is a correlation between hypertension and platelet counts[4]. Also, patients who take antiplatelet drugs can decrease the risk of cardiovascular disease, and patients taking antihypertensive drugs also decrease the risk of cardiovascular disease. Thus, there may be some relationships of platelet count on hypertension. However, if the risk factor has a noncausal association with an outcome, then public health or pharmaceutical interventions targeted at the risk factor will realize no material benefit. Consequently, finding out more potential risk factors of hypertension and establishing the causal relationship is a very emergent public health improvement issue.

Mendelian randomization studies (MR) assess causal inference by using genetic alleles as unbiased proxies for circulating biomarkers. MR studies are based on the random assortment of genetic alleles during meiosis that can confer advantages similar to a randomized controlled trial by investigating the relationship between genetic alleles that are exclusively associated with a biomarker of interest and disease risk[5]. Our study used 16,000 Taiwanese participants selected from the baseline data in the population-based

Taiwan Biobank database with well-designed health and lifestyle and genetic data. We elucidate the causation and reverse causation of platelet count and hypertension with an one-sample setting Mendelian randomization.



Chapter 2 METHOD

2.1 Study population

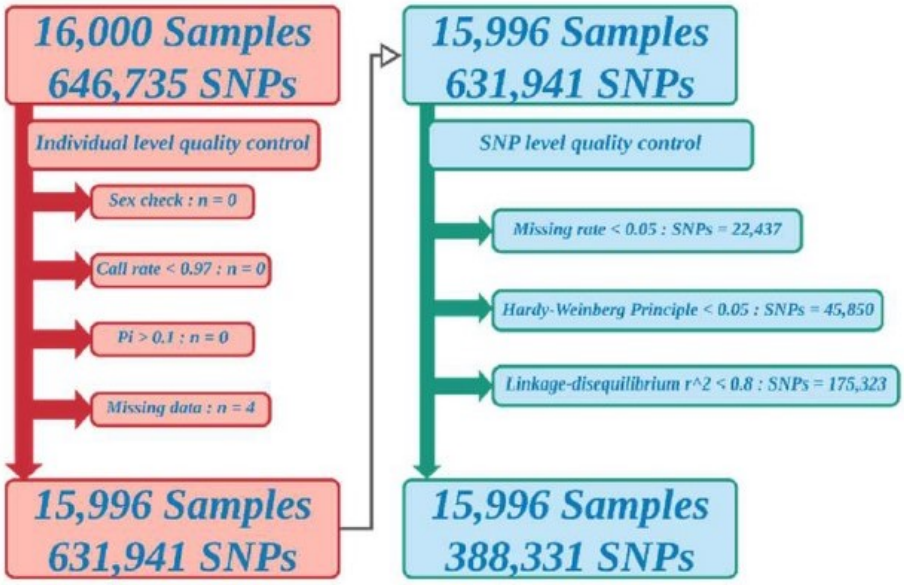
Taiwan Biobank intends to conduct large-scale cohort studies and case-control studies on local diseases by combining genetic and medical information. The community-based cohort study recruits volunteers between 30 and 70 years of age with no history of cancer. The hospital-based cohort study recruits patients affected by the most common chronic diseases in Taiwan, including cardiovascular disease, diabetes, chronic kidney disease, etc. There were 16,000 Taiwanese Han subjects randomly retrieved from the Taiwan Biobank from 2008 to 2015 for conducting the genome-wide study, and these people were taken in our study. Using Axiom-Taiwan Biobank Array Plate (TWB chip; Affymetrix Inc, CA, USA), selected a total of 653,291 gene variant sites and recorded 646,735 single nucleotide polymorphism sites (SNPs).

2.2 Quality control

As Figure 2.1 shows, we have first done the individual quality control and genotyping quality control. At individual quality control, no one was removed because sex mismatch problem among the samples in our study. There was no removal of participants

at a call rate > 0.97. Identity – by-descent was also conducted, and all of the samples passed the cryptic relatedness with pi-hat > 0.1. Nevertheless, we removed four subjects with missing data of platelet count. Thus, 15,996 subjects were included in our study. At genotyping quality control, there were 646,735 SNPs observed in autosome for the SNP-leveled quality control by using PLINK 1.90 beta. Removal of 14,794 variants was carried out at a call rate > 0.97, and 22,437 variants were excluded at the criteria of genotyping missing rate > 0.05. Furthermore, there were 45,850 variants then removed by Hardy-Weinberg tests with p-value < 0.05, and 175,323 variants were pruned by failing to pass linkage-disequilibrium with correlation r < 0.8. There were 15,996 participants and 388,331 variants remaining for the study.

Figure 2.1: Quality control workflow



2.3 Definition of Hypertension and Platelet count

Since our data record from 2008 to 2015, we took the previous definition of American Heart Association (https://www.heart.org/) as a reference to define hypertension as

a dichotomous outcome. Three criteria of inclusion were adopted by an average sitting systolic blood pressure ≥ 140 mmHg, average sitting diastolic blood pressure ≥ 90 mmHg, or self-reported to have hypertension in a questionnaire to decide the hypertensive participants.As Figure 2.2 (a), 21.7% of the analyzed participants were hypertensive. Platelet was selected as the type of platelet count at the baseline measurement in Taiwan Biobank with per unit of $1000/\mu l$. The normal range of platelet count is widely distributed from 150 to 500$(1000/\mu l)$, and the number is susceptible to external change[6]. Figure 2.2 (b) shows the density plot of platelet count stratified by hypertension with all possible confounders unadjusted. Before adjusted any confounders, participants with hypertension have a lower mean platelet count than participants without hypertension.

Figure 2.2: Descriptive analysis of hypertension and platelet count

(a) Pie chart of hypertension (b) Density of platelet count

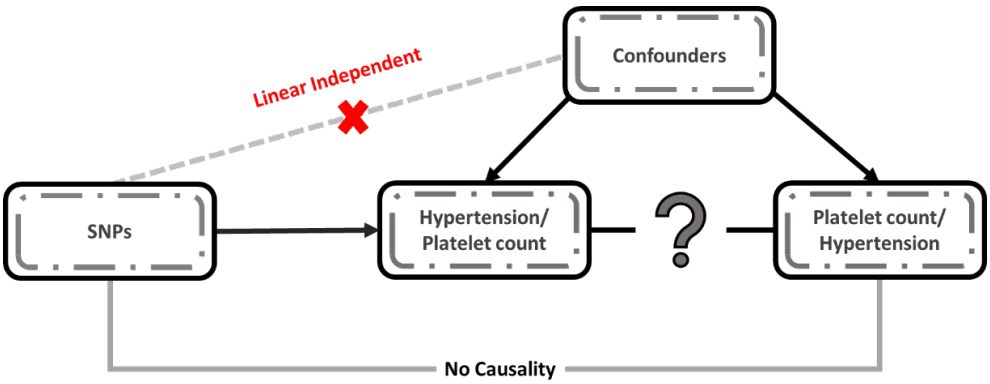
2.4 Association between Hypertension and Platelet count

It has been confirmed that there is a correlation between hypertension and platelet counts in the previous study[4]. Here, we check if the association also exists in our dataset. In reference to previous literature, we took all confounding factors listed below (sex, age, fasting glucose, hematocrit, triglyceride, high-density lipoprotein, hemoglobin, red blood cell, and white blood cell) in the model as covariates to adjust. We applied the logistic regression model with all known confounders to get the effect size of platelet count on hypertension from the model and did Wald test for the significance test. In the reverse direction, we applied multiple linear regression with all the known confounders as covariates and t-value to get the effect size of hypertension on platelet count.

2.5 Mendelian randomization

In order to have a causal relationship between platelet count and hypertension, we applied Mendelian randomization[7] [8]. The causal relationship obtained by genetic variants involved to be as instrumental variables was based on three main assumptions of Mendelian randomization[9]. First, the variant is predictive of the risk factor. Second, the variant is independent of any confounding factors of the risk factor -outcome association. Third, the variant is conditionally independent of the outcome given the risk factor and the confounding factors (Figure 2.2). The selected genetic variants can only affect the outcome via the risk factor if they meet the above conditions [8] [10]. In our study, we applied the inverse variance method (IVW) to elucidate the causality between platelet count and hypertension and conduct the MR-Egger method and the Weighted-median method as a sensitivity analysis.

Figure 2.4: Mendelian randomization

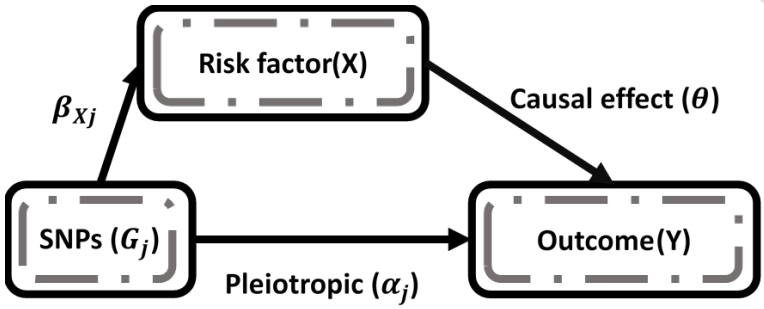


2.5.1 The Framework of data and genetic association

If the association between J genetic variants G_j (j=1,2,⋯,J) and the outcome is denoted β_{Yj} and the association with the risk factor denoted β_{Xj}, then the correlation between the genetic variants and the outcome variable can be expressed as the direct effect

of gene pleiotropy (α_j) plus the indirect causal effect of genetic variants on the outcome through risk factors: $\beta_{Yj} = \alpha_j + \theta\beta_{Xj}$ (Figure 2.3).

Figure 2.5: The magnitude of the causality



2.5.2 IVW method

With a single valid genetic variant G_j, the causal effect of the risk factor on the outcome can be expressed by [8]: $(\hat{\theta}_j) = \frac{\hat{\beta}_{Yj}}{\hat{\beta}_{Xj}}$, where $\hat{\beta}_{Yj}$ indicated the coefficient from univariate regression with the outcome. With multiple genetic variants, the estimates from each genetic variant can be averaged using an inverse-variance weighted (IVW) estimate [11]. This method assumes that the genetic variants are uncorrelated and the pleiotropic effects are zero $\alpha_j = 0$. The regression model can be written as:

$$\hat{\beta}_{Yj} = \theta_{IVW}\hat{\beta}_{Xj} + \epsilon_{ij};\ \epsilon_{ij} \sim \mathcal{N}(0,\ \theta^2 se(\hat{\beta}_{Yj})^2)$$

,where $\hat{\theta}_{IVW} = \frac{\sum_j \hat{\beta}_{Yj}\hat{\beta}_{Xj} se(\hat{\beta}_{Yj})^2}{\sum_j \hat{\beta^2}_{Xj} se(\hat{\beta}_{Yj})^2}$[10] [12]

2.5.3 MR-Egger method

Compare with the IVW method, the MR-Egger method estimates the pleiotropic effects as part of the analysis. The MR-Egger method should conform to the InSIDE assumption (Instrument Strength Independent of Direct effect), which assume that the pleiotropic

effects α_j are independently distributed from the genetic association with the risk factor [13]. The regression model can be written as:

$$\hat{\beta}_{Yj} = \theta_{0E} + \theta_{1E}\hat{\beta}_{Xj} + \epsilon_{Ej};\ \epsilon_{Ej} \sim \mathcal{N}(0,\ \sigma^2 se(\hat{\beta}_{Yj})^2)$$

, where θ_{0E} is the intercept and θ_{1E} is the slope [14].

2.5.4 Median-based method

The median-based methods have greater robustness to individual genetic variants with strongly outlying causal estimates compared with the inverse-variance weighted and MR-Egger methods. Calculate the median of the ratio instrumental variable estimates evaluated using each genetic variant individually. The simple median method gives a consistent estimate of the causal effect when at least 50% of the genetic variants are valid instrumental variables. For the weighted median method, 50% of the weight comes from valid instrumental variables. Also, it will not be affected by outliers and high leverage genetic variants [15].

Chapter 3 RESULTS

According to the definition of the American Heart Association, there were 3,480 hypertensive patients taking account for 21.76% among all participants (Table 3.1). The percentage was a little lower than the previous reports in Taiwan[16] since Taiwan Biobank only included healthy people, and hypertension was known to be the cause of serval diseases. Platelet count with a mean of 237.7 (1000/μL) and standard deviation of 56.9 (1000/μL) was revealed in the analyzed participants. We did the crude association test between platelet count and hypertension in our dataset. Table 3.2 shows that platelet count was significantly positively correlated to hypertension when adjusting sex, age, fasting glucose, hematocrit, triglyceride, high-density lipoprotein, hemoglobin, red blood cell, and white blood cell. With the reverse direction, Table 3.3 shown that hypertension was significantly positively correlated to platelet count when adjusting the same confounders.

Table 3.1: Characteristic of study participants in Taiwan Biobank dataset

Characteristics	**Full sample(N=15996)**
Age(years)	48.7 ± 11.35
Male(sex)	7,965(49.79%)
Hypertension	3,480(21.76%)
Platelet count	237.7 ± 56.92

Continued on next page

Table 3.1 – continued from previous page

Characteristics	Full sample(N=15996)
Fasting glucose	96.5 ± 21.02
Hematocrit	43.7 ± 4.55
Triglyceride	117.5 ± 92.15
High-density lipoprotein	53.1 ± 13.11
Hemoglobin	14.0 ± 1.58
Red blood cell	4.8 ± 0.52
White blood cell	6.1 ± 1.59
†All data are presented as mean ± SD or numbers(%)	

Table 3.2: Logistic regression of Platelet count on Hypertension with all possible confounders adjusted

	Beta	Standard Error	P-value
Platelet count	$1.21e-03$	$5.99-05$	$p<0.05$
Sex(male=0)	$-5.13e-02$	$8.50-03$	$p<0.001$
Age	$1.14e-02$	$2.83-04$	$p<0.001$
Fasting glucose	$1.32-03$	$1.52-04$	$p<0.001)$
Hematocrit	-6.13	$1.42-03$	$p<0.001$
Triglyceride	$1.70-04$	$3.71-05$	$p<0.001$
High−density lipoprotein	$-1.96-03$	$2.72-04$	$p<0.001$
Hemoglobin	$2.07-02$	$4.26-03$	$p<0.001$
Red blood cell	$3.02-02$	$7.47-03$	$p<0.001$
White blood cell	$1.86-02$	$2.09-03$	$p<0.001$

Table 3.3: Linear regression of Hypertension on Platelet count with all possible confounders adjusted

	Beta	**Standard Error**	**P-value**
Hypertension	2.11	1.04	$p < 0.05$
Sex(male=0)	14.02	1.12	$p < 0.001$
Age	-1.02	0.04	$p < 0.001$
Fasting glucose	0.03	0.02	$p = 0.14)$
Hematocrit	-1.67	1.88	$p < 0.001$
Triglyceride	0.03	0.01	$p < 0.001$
High-density lipoprotein	-0.01	0.04	$p = 0.80$
Hemoglobin	-5.09	0.56	$p < 0.001$
Red blood cell	5.70	0.99	$p < 0.001$
White blood cell	10.70	0.26	$p < 0.001$

3.1 The causal effect of Platelet count on Hypertension

Through the genome-wide association studies (GWAS) in recent 10 years with sample size larger than 10,000, as shown in Table A.1, we found 5 single nucleotide polymorphisms (SNPs), which were rs385893 in JAK2, rs11082304 in CABLES1, rs6425521 in DNM3, rs4895441 in HMIP, and rs7775698 in HBS1L, associated with platelet count (Table 3.4). The variants were also significantly associated (p-value < 5e-6) in the Taiwan Biobank dataset. We then performed one-sample Mendelian randomization for the causal inference for platelet on hypertension count by different methods. In Table 3.5, we observed a significant positive casual effect with the simple median, weighted me-

dian and IVW methods. However, there is no significant causal effect with the MR-egger method. Also, the MR-Egger method' s intercept was insignificant. The 5 SNPs related to hypertension did not have a pleiotropic effect.

Table 3.4: Select SNPs correlated with Platelet count as p < 5e-06

SNP	**Gene**	**CHR**	**beta**	**P-value**
rs6425521	*DNM3*	1	4.42	$p < 5e-06$
rs7775698	*HBS1L*	6	8.27	$p < 5e-06$
rs4895441	*HMIP*	6	7.34	$p < 5e-06$
rs385893	*JAK2*	9	-4.99	$p < 5e-06$
rs11082304	*CABLES1*	18	3.2	$p < 5e-06$

†Adjusted all possible confounders, and top 10 principle components from genetic analysis

Table 3.5: Causal estimates of Hypertension on Platelet count

Method	**Estimate**	**Standard Error**	**95%CI**	**P-value**
Simple-median	0.139	0.012	[0.115, 0.162]	$p < 0.05$
Weighted-median	0.134	0.009	[0.116, 0.152]	$p < 0.05$
IVW	0.121	0.061	[0.001, 0.240]	$p < 0.05$
MR-Eggger	0.048	0.208	$[-0.358, 0.455]$	$p = 0.816$
(Intercept)	0.444	1.206	$[-1.919, 2.807]$	$p = 0.713$

3.2 The causal effect of Hypertension on Platelet count

In the reverse direction, through the genome-wide association studies (GWAS) in recent 10 years with sample size larger than 10,000, as shown in Table A.2, we found 6 single nucleotide polymorphisms (SNPs), which were rs1458038 in FGF5, rs3796605 in FGF5, rs455938 in MAST4, rs10866754 in CTC-535M15.2, rs648435 in APHGAP42, and rs2018159 in APHGAP42 significantly associated (p-value < 5e-6) with hypertension in Taiwan Biobank dataset (Table 3.6), . We performed one-sample Mendelian randomization for the causal inference for hypertension on platelet count by different methods. In Table 3.7, no significant causal effect of hypertension on platelet count was observed. Since there is no significant effect of the MR-Egger method' s intercept, the 6 SNPs related to hypertension did not have a pleiotropic effect. Both IVW and Weighted methods indicate that the weighted causal effect is positive but not significant.

Table 3.6: Select SNPs correlated with Platelet count as p < 5e-06

SNP	**Gene**	**CHR**	**beta**	**P-value**
rs1458038	*FGF5*	4	1.197	$p < 5e-06$
rs3796605	*FGF5*	4	0.8562	$p < 5e-06$
rs455938	*MAST4*	5	1.15	$p < 5e-06$
rs10866754	*CTC-535M15.2*	1.169	-4.99	$p < 5e-06$
rs648435	*APHGAP42*	11	0.8616	$p < 5e-06$
rs2018159	*APHGAP42*	11	0.8563	$p < 5e-06$

†Adjusted all possible confounders, and top 10 principle components from genetic analysis

Table 3.7: Causal estimates of Hypertension on Platelet count

Method	**Estimate**	**Standard Error**	**95%CI**	**P-value**
Simple-median	0.446	0.315	$[-1.171,\ 1.063]$	$p = 0.156$
Weighted-median	0.254	0.316	$[-0.366,\ 0.874]$	$p = 0.423$
IVW	0.343	0.258	$[-0.164,\ 0.849]$	$p = 0.185$
MR-Eggger	-1.294	0.694	$[-4.613,\ 2.025]$	$p = 0.445$
(Intercept)	1.703	1.714	$[-1.710, 5.166]$	$p = 0.328$

Chapter 4 DISCUSSION

4.1 Mendelian randomization assumptions

With Mendelian randomization assumptions, only the first assumption can be fully empirically tested because second and third assumptions depend on all possible confounders of risk factor-outcome association, both measured and unmeasured. While using the IVW method, we need all genetic variants to satisfy the MR assumptions to elucidate a consistent estimate of the causal effect[15]. Hence, we conduct the MR-Egger method and the Weighted-median method as a sensitivity analysis. The MR-Egger method estimates the true causal effect that is consistent even if all genetic are invalid due to violation of the third assumption but under a weaker assumption is known as InSIDE (instrument strength independent of direct effect) assumption[13]. However, MR-Egger regression estimates are likely to be particularly imprecise if all genetic variants have similar magnitudes of association with the risk factor. The weighted median method will provide a consistent estimate if at least 50% of the weight comes from valid genetic variants and assume that no single genetic variant contributes more than 50% of the weighted. Compare with the MR-Egger method, the weighted median method approach allows the MR assumptions to be violated in a more general way for the invalid genetic variants[15]. Consequently, although we observed an insignificant estimate in the MR-Egger method, we believe that

there is a causal effect of platelet count on hypertension.

4.2 Limitations

In our study, some potential limitations exist. First, we had only baseline platelet count in the Taiwan Biobank. These limitations would arise less robustness due to the individual variability and should be exchanged by the average of the measurements several repeated times. Second, the definition of hypertension included self-reported to have hypertension in the questionnaire. Recall bias was inevitable when the questionnaire was used and then misclassified some of the hypertensive patients. Due to the misclassification of the hypertensive patients, we may underestimated the magnitude of the effect. Third, the data we used in the Taiwan Biobank dataset is a cross-sectional study. Due to the data restriction, we could only observe one direction of the causal effect simultaneously, although reverse causation exists. Last, as for the method, we need to adjust all possible confounders to comply with the Mendelian randomization assumptions. However, there were still existing some unknown confounding factors. The causality was not guaranteed if there were unknown confounding factors in the relationship. The result could be impacted by the limitations, which should be cautious of when making further applications.

Chapter 5 CONCLUSION

In our study, we revealed the significant positive causal relationship of platelet count on hypertension by Mendelian randomization. However, there is no significant causal effect in the reverse direction. Platelet count can be taken as one of the risk factors of hypertension, provide the evaluation reference for potential hypertension in clinical diagnosis, and can set the stricter threshold of hypertension to keep track of for those who have higher platelet count to prevent hypertension. Furthermore, combined with the relationships of other platelet indices on hypertension under larger and multiple data sources in future studies, we can have more evidence on platelet and hypertension to develop more therapeutic treatments on hypertension.

References

[1] K. R. Czuriga-Kovács, D. Czuriga, and L. Csiba. Influence of hypertension, alone and in combination with other vascular risk factors on cognition. CNS Neurol Disord Drug Targets, 15(6):690–8, 2016.

[2] M. R. Lee, Y. H. Lim, and Y. C. Hong. Causal association of body mass index with hypertension using a mendelian randomization design. Medicine (Baltimore), 97(30):e11252, 2018.

[3] S. K. Kunutsor, L. M. Kieneker, S. Burgess, S. J. L. Bakker, and R. P. F. Dullaart. Circulating total bilirubin and future risk of hypertension in the general population: The prevention of renal and vascular end-stage disease (prevend) prospective study and a mendelian randomization approach. J Am Heart Assoc, 6(11), 2017.

[4] Jianan Zhang, Jing Li, Shi Chen, Linglin Gao, Xiaoluan Yan, Mingzhi Zhang, Jia Yu, Fenchun Wang, and Hao Peng. Modification of platelet count on the association between homocysteine and blood pressure: A moderation analysis in chinese hypertensive patients. International Journal of Hypertension, 2020:1–8, 2020.

[5] George Davey Smith and Shah Ebrahim. What can mendelian randomisation tell us about modifiable behavioural and environmental exposures? BMJ, 330(7499):1076–1079, 2005.

[6] R. Sachdev, A. K. Tiwari, S. Goel, V. Raina, and M. Sethi. Establishing biological reference intervals for novel platelet parameters (immature platelet fraction, high immature platelet fraction, platelet distribution width, platelet large cell ratio, platelet-x, plateletcrit, and platelet distribution width) and their correlations among each other. Indian J Pathol Microbiol, 57(2):231–5, 2014.

[7] George Davey Smith and Shah Ebrahim. ‘mendelian randomization’: can genetic epidemiology contribute to understanding environmental determinants of disease?*. International Journal of Epidemiology, 32(1):1–22, 2003.

[8] Debbie A. Lawlor, Roger M. Harbord, Jonathan A. C. Sterne, Nic Timpson, and George Davey Smith. Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. Statistics in Medicine, 27(8):1133–1163, 2008.

[9] E. P. Martens, W. R. Pestman, A. de Boer, S. V. Belitser, and O. H. Klungel. Instrumental variables: application and limitations. Epidemiology, 17(3):260–7, 2006.

[10] V. Didelez and N. Sheehan. Mendelian randomization as an instrumental variable approach to causal inference. Stat Methods Med Res, 16(4):309–30, 2007.

[11] S. Burgess, A. Butterworth, and S. G. Thompson. Mendelian randomization analysis with multiple genetic variants using summarized data. Genet Epidemiol, 37(7):658–65, 2013.

[12] G. Rücker, G. Schwarzer, J. R. Carpenter, H. Binder, and M. Schumacher. Treatment-effect estimates adjusted for small-study effects via a limit meta-analysis. Biostatistics, 12(1):122–42, 2011.

[13] Michal Kolesár, Raj Chetty, John Friedman, Edward Glaeser, and Guido W. Imbens. Identification and inference with many invalid instruments. Journal of Business Economic Statistics, 33(4):474–484, 2015.

[14] Stephen Burgess and Simon G. Thompson. Use of allele scores as instrumental variables for mendelian randomization. International Journal of Epidemiology, 42(4):1134–1144, 2013.

[15] J. Bowden, G. Davey Smith, P. C. Haycock, and S. Burgess. Consistent estimation in mendelian randomization with some invalid instruments using a weighted median estimator. Genet Epidemiol, 40(4):304–14, 2016.

[16] P. M. Kearney, M. Whelton, K. Reynolds, P. Muntner, P. K. Whelton, and J. He. Global burden of hypertension: analysis of worldwide data. Lancet, 365(9455):217–23, 2005.

[17] Masahiro Kanai, Masato Akiyama, Atsushi Takahashi, Nana Matoba, Yukihide Momozawa, Masashi Ikeda, Nakao Iwata, Shiro Ikegawa, Makoto Hirata, Koichi Matsuda, Michiaki Kubo, Yukinori Okada, and Yoichiro Kamatani. Genetic analysis of quantitative traits in the japanese population links cell types to complex human diseases. Nature Genetics, 50(3):390–400, 2018.

[18] W. J. Astle, H. Elding, T. Jiang, D. Allen, D. Ruklisa, A. L. Mann, D. Mead, H. Bouman, F. Riveros-Mckay, M. A. Kostadima, J. J. Lambourne, S. Sivapalaratnam, K. Downes, K. Kundu, L. Bomba, K. Berentsen, J. R. Bradley, L. C. Daugherty, O. Delaneau, K. Freson, S. F. Garner, L. Grassi, J. Guerrero, M. Haimel, E. M. Janssen-Megens, A. Kaan, M. Kamat, B. Kim, A. Mandoli, J. Marchini, J. H. A. Martens, S. Meacham, K. Megy, J. O'Connell, R. Petersen, N. Sharifi, S. M. Sheard,

J. R. Staley, S. Tuna, M. van der Ent, K. Walter, S. Y. Wang, E. Wheeler, S. P. Wilder, V. Iotchkova, C. Moore, J. Sambrook, H. G. Stunnenberg, E. Di Angelantonio, S. Kaptoge, T. W. Kuijpers, E. Carrillo-de Santa-Pau, D. Juan, D. Rico, A. Valencia, L. Chen, B. Ge, L. Vasquez, T. Kwan, D. Garrido-Martín, S. Watt, Y. Yang, R. Guigo, S. Beck, D. S. Paul, T. Pastinen, D. Bujold, G. Bourque, M. Frontini, J. Danesh, D. J. Roberts, W. H. Ouwehand, A. S. Butterworth, and N. Soranzo. The allelic landscape of human blood cell trait variation and links to common complex disease. Cell, 167(5):1415–1429.e19, 2016.

[19] U. M. Schick, D. Jain, C. J. Hodonsky, J. V. Morrison, J. P. Davis, L. Brown, T. Sofer, M. P. Conomos, C. Schurmann, C. P. McHugh, S. C. Nelson, S. Vadlamudi, A. Stilp, A. Plantinga, L. Baier, S. A. Bien, S. M. Gogarten, C. A. Laurie, K. D. Taylor, Y. Liu, P. L. Auer, N. Franceschini, A. Szpiro, K. Rice, K. F. Kerr, J. I. Rotter, R. L. Hanson, G. Papanicolaou, S. S. Rich, R. J. Loos, B. L. Browning, S. R. Browning, B. S. Weir, C. C. Laurie, K. L. Mohlke, K. E. North, T. A. Thornton, and A. P. Reiner. Genome-wide association study of platelet count identifies ancestry-specific loci in hispanic/latino americans. Am J Hum Genet, 98(2):229–42, 2016.

[20] J. H. Oh, Y. K. Kim, S. Moon, Y. J. Kim, and B. J. Kim. Genome-wide association study identifies candidate loci associated with platelet count in koreans. Genomics Inform, 12(4):225–30, 2014.

[21] A. Gusev, S. H. Lee, G. Trynka, H. Finucane, B. J. Vilhjálmsson, H. Xu, C. Zang, S. Ripke, B. Bulik-Sullivan, E. Stahl, A. K. Kähler, C. M. Hultman, S. M. Purcell, S. A. McCarroll, M. Daly, B. Pasaniuc, P. F. Sullivan, B. M. Neale, N. R. Wray, S. Raychaudhuri, and A. L. Price. Partitioning heritability of regulatory and cell-

type-specific variants across 11 common diseases. Am J Hum Genet, 95(5):535–52, 2014.

[22] C. Gieger, A. Radhakrishnan, A. Cvejic, W. Tang, E. Porcu, G. Pistis, J. Serbanovic-Canic, U. Elling, A. H. Goodall, Y. Labrune, L. M. Lopez, R. Mägi, S. Meacham, Y. Okada, N. Pirastu, R. Sorice, A. Teumer, K. Voss, W. Zhang, R. Ramirez-Solis, J. C. Bis, D. Ellinghaus, M. Gögele, J. J. Hottenga, C. Langenberg, P. Kovacs, P. F. O'Reilly, S. Y. Shin, T. Esko, J. Hartiala, S. Kanoni, F. Murgia, A. Parsa, J. Stephens, P. van der Harst, C. Ellen van der Schoot, H. Allayee, A. Attwood, B. Balkau, F. Bastardot, S. Basu, S. E. Baumeister, G. Biino, L. Bomba, A. Bonnefond, F. Cambien, J. C. Chambers, F. Cucca, P. D'Adamo, G. Davies, R. A. de Boer, E. J. de Geus, A. Döring, P. Elliott, J. Erdmann, D. M. Evans, M. Falchi, W. Feng, A. R. Folsom, I. H. Frazer, Q. D. Gibson, N. L. Glazer, C. Hammond, A. L. Hartikainen, S. R. Heckbert, C. Hengstenberg, M. Hersch, T. Illig, R. J. Loos, J. Jolley, K. T. Khaw, B. Kühnel, M. C. Kyrtsonis, V. Lagou, H. Lloyd-Jones, T. Lumley, M. Mangino, A. Maschio, I. Mateo Leach, B. McKnight, Y. Memari, B. D. Mitchell, G. W. Montgomery, Y. Nakamura, M. Nauck, G. Navis, U. Nöthlings, I. M. Nolte, D. J. Porteous, A. Pouta, P. P. Pramstaller, J. Pullat, S. M. Ring, J. I. Rotter, D. Ruggiero, A. Ruokonen, C. Sala, N. J. Samani, J. Sambrook, D. Schlessinger, et al. New gene functions in megakaryopoiesis and platelet formation. Nature, 480(7376):201–8, 2011.

[23] Y. Kamatani, K. Matsuda, Y. Okada, M. Kubo, N. Hosono, Y. Daigo, Y. Nakamura, and N. Kamatani. Genome-wide association study of hematological and biochemical traits in a japanese population. Nat Genet, 42(3):210–5, 2010.

[24] Daniel Levy, Georg B. Ehret, Kenneth Rice, Germaine C. Verwoert, Lenore J.

Launer, Abbas Dehghan, Nicole L. Glazer, Alanna C. Morrison, Andrew D. Johnson, Thor Aspelund, Yurii Aulchenko, Thomas Lumley, Anna Köttgen, Ramachandran S. Vasan, Fernando Rivadeneira, Gudny Eiriksdottir, Xiuqing Guo, Dan E. Arking, Gary F. Mitchell, Francesco U. S. Mattace-Raso, Albert V. Smith, Kent Taylor, Robert B. Scharpf, Shih-Jen Hwang, Eric J. G. Sijbrands, Joshua Bis, Tamara B. Harris, Santhi K. Ganesh, Christopher J. O'Donnell, Albert Hofman, Jerome I. Rotter, Josef Coresh, Emelia J. Benjamin, André G. Uitterlinden, Gerardo Heiss, Caroline S. Fox, Jacqueline C. M. Witteman, Eric Boerwinkle, Thomas J. Wang, Vilmundur Gudnason, Martin G. Larson, Aravinda Chakravarti, Bruce M. Psaty, and Cornelia M. Van Duijn. Genome-wide association study of blood pressure and hypertension. Nature Genetics, 41(6):677–687, 2009.

[25] Xiangfeng Lu, Laiyuan Wang, Xu Lin, Jianfeng Huang, C. Charles Gu, Meian He, Hongbing Shen, Jiang He, Jingwen Zhu, Huaixing Li, James E. Hixson, Tangchun Wu, Juncheng Dai, Ling Lu, Chong Shen, Shufeng Chen, Lin He, Zengnan Mo, Yongchen Hao, Xingbo Mo, Xueli Yang, Jianxin Li, Jie Cao, Jichun Chen, Zhongjie Fan, Ying Li, Liancheng Zhao, Hongfan Li, Fanghong Lu, Cailiang Yao, Lin Yu, Lihua Xu, Jianjun Mu, Xianping Wu, Ying Deng, Dongsheng Hu, Weidong Zhang, Xu Ji, Dongshuang Guo, Zhirong Guo, Zhengyuan Zhou, Zili Yang, Renping Wang, Jun Yang, Xiaoyang Zhou, Weili Yan, Ningling Sun, Pingjin Gao, and Dongfeng Gu. Genome-wide association study in chinese identifies novel loci for blood pressure and hypertension. Human Molecular Genetics, 24(3):865–874, 2015.

[26] Xin Liu, Cheng Hu, Minghui Bao, Jing Li, Xiaoyan Liu, Xuerui Tan, Yong Zhou, Yequn Chen, Shouling Wu, Shuohua Chen, Rong Zhang, Feng Jiang, Weiping Jia, Xingyu Wang, Xinchun Yang, and Jun Cai. Genome wide association study iden-

tifies l3mbtl4 as a novel susceptibility gene for hypertension. Scientific Reports, 6(1):30811, 2016.

[27] P. Surendran, F. Drenos, R. Young, H. Warren, J. P. Cook, A. K. Manning, N. Grarup, X. Sim, D. R. Barnes, K. Witkowska, J. R. Staley, V. Tragante, T. Tukiainen, H. Yaghootkar, N. Masca, D. F. Freitag, T. Ferreira, O. Giannakopoulou, A. Tinker, M. Harakalova, E. Mihailov, C. Liu, A. T. Kraja, S. Fallgaard Nielsen, A. Rasheed, M. Samuel, W. Zhao, L. L. Bonnycastle, A. U. Jackson, N. Narisu, A. J. Swift, L. Southam, J. Marten, J. R. Huyghe, A. Stančáková, C. Fava, T. Ohlsson, A. Matchan, K. E. Stirrups, J. Bork-Jensen, A. P. Gjesing, J. Kontto, M. Perola, S. Shaw-Hawkins, A. S. Havulinna, H. Zhang, L. A. Donnelly, C. J. Groves, N. W. Rayner, M. J. Neville, N. R. Robertson, A. M. Yiorkas, K. H. Herzig, E. Kajantie, W. Zhang, S. M. Willems, L. Lannfelt, G. Malerba, N. Soranzo, E. Trabetti, N. Verweij, E. Evangelou, A. Moayyeri, A. C. Vergnaud, C. P. Nelson, A. Poveda, T. V. Varga, M. Caslake, A. J. de Craen, S. Trompet, J. Luan, R. A. Scott, S. E. Harris, D. C. Liewald, R. Marioni, C. Menni, A. E. Farmaki, G. Hallmans, F. Renström, J. E. Huffman, M. Hassinen, S. Burgess, R. S. Vasan, J. F. Felix, M. Uria-Nickelsen, A. Malarstig, D. F. Reily, M. Hoek, T. Vogt, H. Lin, W. Lieb, M. Traylor, H. F. Markus, H. M. Highland, A. E. Justice, E. Marouli, J. Lindström, M. Uusitupa, P. Komulainen, T. A. Lakka, et al. Trans-ancestry meta-analyses identify rare and common variants associated with blood pressure and hypertension. Nat Genet, 48(10):1151–1161, 2016.

Appendix A — Literature review

Table A.1: genome-wide association studies with genetic variants related to platelet count

paper ID	overlapped gene	publication date	sample size	sample ethics
GWAS in Japan[17]	*JAK2* *CABLES1*	2018-02-05	n=108,208	East Asian
The Allelic Landscape of Human Blood Cell[18]	*CABLES1* *DNM3*	2016-11-17	n=166,066	European
GWAS of platelet in Hispanic or Latin American[19]	*JAK2* *HMIP*	2016-01-21	n=12,491	Hispanic or Latin American
Gwas in Korean[20]	*NA*	2014-12-31	n=8,842	East Asian
GWAS of platelet related[21]	*CABLES1*	2013-09-12	n=13,582	European
New gene function in platelet formation[22]	*CABLES1*	2011-11-30	n=48,666	European
GWAS in Japanese biochemical traits [23]	*CABLES1* *HBS1L*	2012-02-07	n=14,806	East Asian

Table A.2: genome-wide association studies with genetic variants related to hypertension

paper ID	**overlapped gene**	**publication date**	**sample size**	**sample ethics**
GWAS study of blood pressure and hypertension [24]	*ATP2B1* *CASZ1* *CYP17A1* *SH2B3*	2009-05-10	n=29,136	European
GWAS study in Chinese identifies nuvel loci for blood pressure and hypertenison [25]	*ATP2B1* *CASZ1* *FGF5* *CYP17A1*	2014-09-23	n=11,816	Chinese population
GWAS study identifies L3MBTL4 as a Novel Susceptibility Gene for Hypertension [26]	*ATP2B1* *CASZ1* *FGF5* *CYPA1*	2016-08-02	n=16,870	Chunese population
Trans-ancestry meta-analysis identify rare and common variants associated with blood pressure and hypertension [27]	*ATP2B1* *CASZ1* *FGF5*	2016-10-01	n=165,276 n = 192,763	Europeans South Asians