**Title:** Using Multivariable Mendelian Randomisation to determine the causal relationships of cardiovascular risk factors on Heart Failure development.


## Abstract:

Heart Failure (HF) is a complex clinical condition that arises from functional and structural myocardial defects, leading to impaired ventricular filling and ejection of blood. HF has a multifactorial aetiology, making it challenging to fully understand the pathophysiology of this condition. Determining causal relationships of potential risk factors can be a struggle using traditional observational studies due to them being prone to confounding variables and reverse causation. Mendelian Randomization (MR) is the use of genetic variants as instrumental variables for modifiable risk factors to determine their causal effect. MR studies are less susceptible to confounding variables and reverse causation in comparison to observational studies. This study aimed to use Multivariable Mendelian Randomization to investigate the direct causal effects of LDL Cholesterol, Coronary Artery Disease (CAD), Body Mass Index (BMI), Type 2 Diabetes Mellitus, and Systolic blood pressure on the development of HF. We conducted a Two-Sample MVMR analysis using summary statistics with a European ancestry from genome-wide association studies (GWAS). Univariable MR was conducted as a foundation prior to Multivariable MR to account for potential confounding and pleiotropy. Finally, sensitivity analyses such as pleiotropy (MR-Egger intercept) and heterogeneity testing were performed to determine the robustness of our results. Initial Univariable MR results show significant associations between each risk factor and the outcome Heart Failure. Sensitivity analyses displayed no significant pleiotropy being present for each risk factor, however, significant heterogeneity was shown to be present for every exposure. Robust inverse-variance weighted estimates were then calculated to adjust for heterogeneity and outliers, these estimates were significant and not much different from the initial IVW estimates. MVMR was then conducted to estimate the direct causal effects of each exposure, and a phenotypic correlation matrix was used to adjust for correlation as it is unlikely our data was from independent samples. Body mass index, Systolic blood pressure, and coronary artery disease all showed significant positive associations with Heart Failure, however, these results were not robust to pleiotropic bias due to the presence of significant heterogeneity. Robust estimates were then calculated by minimising the Q-Statistic value. Significant decreases in the causal estimates of systolic blood pressure and coronary artery disease were seen and the causal estimates for both LDL Cholesterol and Type 2 diabetes had increased. However, under close inspection, the coronary artery disease summary statistics included participants of non-European ancestry (Japanese). Therefore, bias due to population stratification would be present within the analysis and would need to be considered. In conclusion, MVMR results displayed significant causal relationships between coronary artery disease, BMI, and systolic blood pressure with Heart Failure. These exposures could be targets for the development of future interventions for heart failure prevention. However, these results may be subject to potential biases due to population stratification caused by the coronary artery disease dataset.


## Introduction:

Heart Failure (HF) is a clinical syndrome that is caused by functional and structural defects within the myocardium. This results in the impairment of blood ejection as well as ventricular filling [1]. The main types of heart failure are left-sided, right-sided, and biventricular (both sides) heart failure. The most common type is left-sided heart failure which tends to cause breathing symptoms and can be classified into heart failure with reduced or preserved ejection fraction. Heart failure with reduced ejection fraction (HFrEF) is when the left ventricle is weakened and is incapable of contracting normally, this results in a reduced ejection fraction (amount of blood pumped out with each contraction), typically less than 40% [2]. On the other hand, the ejection fraction for heart failure with preserved ejection fraction (HFpEF) is normal but the left ventricle is stiff and does not fill properly between each heartbeat, HFpEF is more common in older individuals or people with high blood pressure[3].

HF is a highly prevalent ( > 37.7 million individuals globally (2016)) disorder whose mechanisms are not completely understood, this is due to the multifactorial nature and heterogeneity of this disease[4]. Observational studies have been used to analyse associations between risk factors and heart failures by tracking individuals over a period of time[5]. However, these studies are prone to be limited by confounding

variables as well as other factors such as reverse causation. An example can be an observational study that examines the association between high blood pressure and heart failure. The age of an individual can be a confounding variable, and due to the risk of heart failure and high blood pressure naturally increasing with age, if this confounding factor is left unaccounted for, it can result in unreliable conclusions of high blood pressure directly causing heart failure more than it truly does. Mendelian Randomisation (MR) is a statistical technique that allows for the estimation of the causal effect of selected risk factors on an outcome of interest whilst being able to adjust for confounding variables and other factors that may impact the true causal effect[6]. These statistical techniques could allow us to develop a further understanding of the pathophysiology of heart failure which can help the production of future clinical interventions for this condition.

### Fundamentals of MR:

Mendelian randomization (MR) is the use of genetic variants as instrumental variables for modifiable risk factors to determine their causal effect. Instrumental variables are variables that have an association with the risk factors of interest, these are not related to any confounders and only affect the outcome through the risk factor itself. MR can be used to address the key challenges of confounding variables and reverse causation within observational studies[7]. Confounding factors are variables that are associated with both the exposure and outcome, and these can distort the estimated causal effect if they are not properly accounted for. Reverse causation is when the outcome is shown to influence the exposure/risk factor and not vice versa. The presence of confounding factors and reverse causation can result in bias within the causal effect, and it can also affect the validity of the study[8]. MR studies are less susceptible to confounding variables in comparison to observational studies due to genetic variants being randomly allocated and generally not being influenced by confounding environmental and lifestyle factors according to Mendel's laws. Furthermore, the risk of reverse causation being present in MR studies is not likely due to genetic variants causing the onset of the outcome whilst remaining unaffected by it. Finally, MR can be conducted on both summary and individual data from genome-wide association studies (GWAS) to determine the causality of selected exposures on an outcome.

### Univariable Mendelian Randomisation and its Assumptions:

Univariable MR is the use of genetic variants as instrumental variables to determine the total causal effect of a single exposure on a single outcome. There are three assumptions that must be true for a Mendelian randomization study to be valid. First the Relevance assumption (IV1), the genetic variants must be associated with the risk factor of interest, secondly, the independence assumption (IV2), where the SNP or genetic variant is not associated with confounding variables. The final assumption is exclusion restriction (IV3), meaning the genetic variant is only associated with the outcome through the exposure/risk factor of interest[9]. Figure 1 displays these three assumptions as well as the associations that should be present in an MR study.

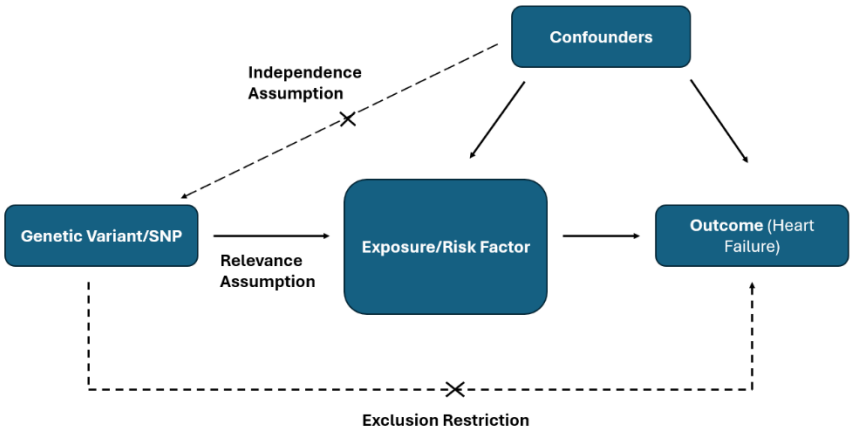


**Figure 1:** Acyclic graph of the three assumptions in Mendelian Randomisation for the outcome Heart Failure. The dashed lines display associations that should not exist for genetic variants to be used as valid instrumental variables for exposures.

## Pleiotropy:

A major obstacle in Mendelian Randomization is the presence of pleiotropy. Pleiotropy is the association of an SNP/Genetic Variant with numerous phenotypes, and this can take two different forms, Vertical and Horizontal pleiotropy. During Vertical Pleiotropy, the genetic variant is associated with another factor throughout the pathway from the variant via the risk factor to the outcome. The causal pathway will always be through the risk factor of interest and this type of pleiotropy does not distort or invalidate any of the causal estimates. Discovering these potential mediators can improve our understanding of the causal pathway and the pathophysiology of disease. Horizontal pleiotropy occurs when the genetic variant also affects the outcome through a different factor that is not related to the exposure of interest, and this creates a confounding effect[10]. This violates the third MR assumption (exclusion restriction) due to the genetic variant affecting the outcome through pathways that are independent of the exposure of interest. Horizontal pleiotropy can lead to false positive findings if the pleiotropic effect of the upstream factor counteracts the true causal effect of the risk factor on the outcome[11]. These pathways can be seen in Figure 2.

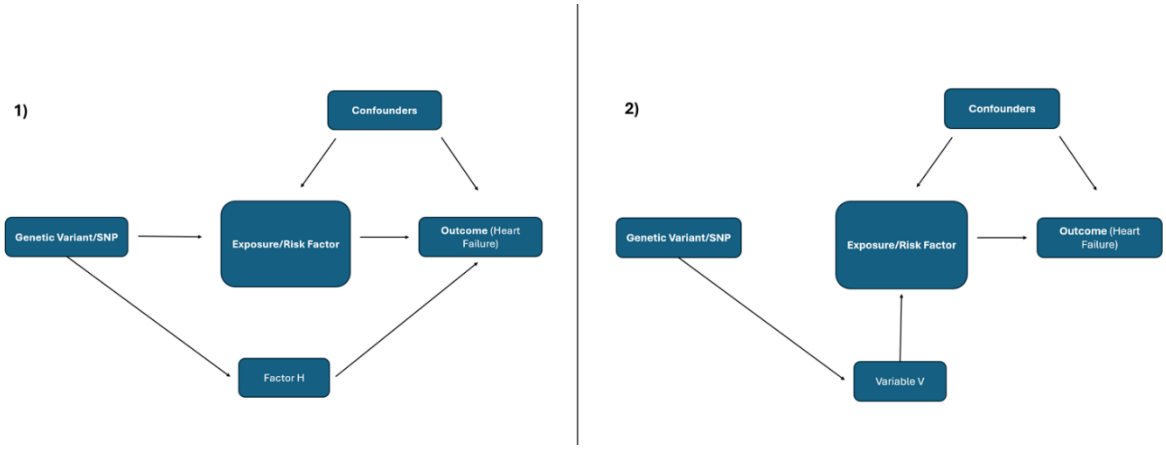



**Figure 2:** Displays 1) Horizontal Pleiotropy which shows Factor H having a causal effect on the outcome which is independent of the exposure/risk factor of interest. 2) Vertical Pleiotropy shows the variant interacting with another factor (Variable V) whilst still having a causal effect on the outcome through the exposure of interest. From Figure 1, you can see that Horizontal Pleiotropy violates the exclusion restriction assumption thus why it is a problem from MR analysis.

## Multivariable Mendelian Randomisation (MVMR):

Univariable MR relies on the use of genetic variants that are associated with a single risk factor, but many variants are shown to be associated with numerous risk factors (pleiotropic). At times, there may be no variants that are solely associated with one risk factor and this analysis cannot be conducted without the consideration of pleiotropic variants. A more powerful analysis can be done by including pleiotropic variants within a study. An advanced extension of MR called Multivariable Mendelian Randomisation allows the estimation of the direct causal effect of numerous exposures (may also be related) on a single outcome, conditional on the included exposures in the model, simultaneously[12]. Thus, MVMR is useful when two potentially related exposures are of interest, and you would like to identify whether both exert a causal effect on the outcome or if one exposure acts as a mediator for the other exposure. The basic model of an MVMR analysis can be seen in Figure 3, this uses two risk factors and the outcome heart failure. The genetic variant is shown to be associated with at least one of the exposures, the line between exposures 1 and 2 is left bidirectional because no assumption is made about their relationship when estimating the direct causal effects of exposure 1 (β1) and exposure 2 (β2) on the outcome. Confounders can be shown to affect the outcome directly as well as both the exposures in the model.

### MVMR Assumptions:

Similarly to normal/univariable Mendelian randomization, MVMR has a set of assumptions for the genetic variants, to ensure the instrument's validity for the conducted study[13]. The three assumptions are as follows:

- **MV-IV1:** Variant must be strongly associated with each exposure given the rest of the exposures within the model.
- **MV-IV2**: Variant is independent of confounders of all exposures and the outcome
- **MV-IV3:** Variant is independent of the outcome given all the exposures

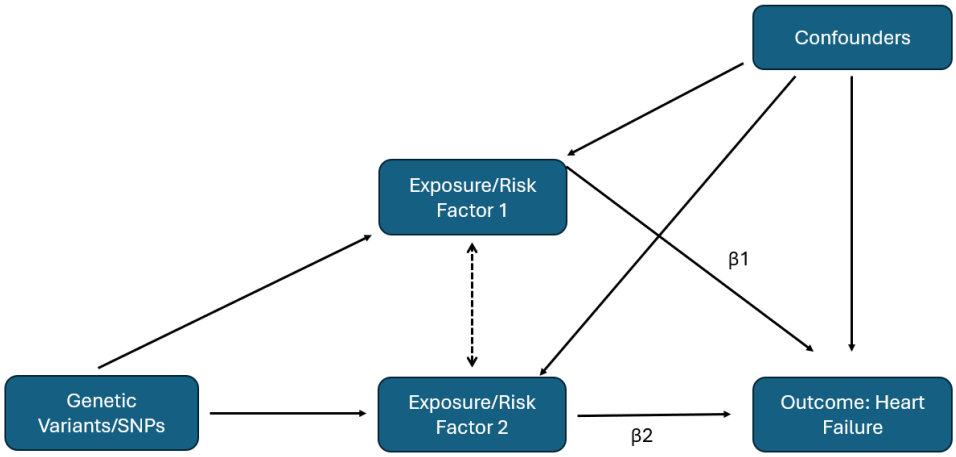



**Figure 3:** Displays an MVMR model for two exposures. The SNPs would be associated with at least one of the exposures. β1 and β2 represent the direct causal effects of Exposure/Risk Factors 1 and 2 on the outcome (Heart Failure), respectively.

### One and Two-Sample MR:

The purpose of One-Sample and Two-Sample MR is to investigate causal relationships between exposures and outcomes of interest whilst using genetic variants as instrumental variables, however, their approaches are slightly different. For One-sample MR, the genetic variant-exposure and genetic variant-outcome associations are estimated using an individual-level dataset, furthermore, the same data is used for both the exposure and outcome. It can also allow the exploration of the associations between the instrument with the confounders observed in the exposure-outcome association, as well as the interactions and sub-group effects (the effect of exposure on outcome may be different due to environmental factors). One-sample MR is more prone to data overfitting, and weak instrument bias will tend to cause the estimate to be biased towards the MV regression estimate[14]. This means if the MV regression displays a positive association between the exposure and outcome, the one-sample MR estimate will tend to be positive even if there is not a true causal effect. On the other hand, for an MV regression result displaying a negative association, the estimate will tend to be negative.

Two-sample MR uses summary data from existing GWAS (Genome-Wide Association Studies), the exposure and outcome data come from different samples[15]. Two-sample MR is considered more powerful as it can use larger datasets and is less susceptible to confounding, due to reasons such as the separation of exposure and outcome samples as well as reduced population-specific biases. This form of MR is less susceptible to data overfitting in comparison to One-Sample MR, and the weak selection instrument bias will tend toward the null. However, if there is an overlap between the individuals in Two-sample MR, the weak instrument bias will tend

the estimate towards the MV regression estimate (like one-sample MR) in proportion to the amount of overlap.

There are several techniques used for two-sample MR to analyse causal relationships such as MR-Egger, inverse-Variance Weighted (IVW), weighted median, and weighted mode. MR-egger regression analysis combines the two sets of data and accounts for potential directional pleiotropy, and the slope of the line indicates the estimated causal effect of the exposure on the outcome whilst accounting for directional pleiotropy. The intercept of the regression indicates the presence of horizontal pleiotropy, with a non-zero intercept suggesting that bias in the results due to directional pleiotropic effects. Another method is IVW, which calculates the weighted average of the effect for each variant, with the weights being inversely proportional to each estimate's variance. This ensures that the variants with a smaller variance (more precise estimate) have a larger impact on the combined estimate, whilst those estimates with a large estimate have a smaller impact on the estimate[16]. The weighted median estimator can be used to provide a consistent estimate of causal effect if at least 50% of the instrumental variables are valid. Each estimate is assigned a weight which is then standardized (sum is equal to 1). The weighted median is then calculated, and we assume that no single IV contributes more than 50% of its weight. if more than 50% of the weight is from valid IVs, the weighted median estimate can be seen as a consistent estimate for the causal effect of the exposure[17]. Thus, this approach is more likely to give a valid causal estimate in comparison to IVW and MR-Egger. The final method is weighted model-based estimation, which assumes the most common causal effect is consistent with the true causal effect, this method has low bias but also has low power to determine a causal effect. One-sample MR would analyse the association between the variant, exposure, and outcome using statistical methods. Most commonly, Two-Stage Least squares (2SLS) is a robust technique that uses linear regression to estimate causal effects through addressing bias and isolating the causal effect.

## Recent MR Techniques:

Causal inference within genetics has been shown to face challenges due to the presence of pleiotropy and traditional MR techniques have been shown to be affected by this. To address this, Jingshu Wang in an article published in 2021 proposed a new framework called GRAPPLE ("Genome-wide mR Analysis under Pervasive PLEIotropy") for analysing the causal effect of risk factors with the use of heterogeneous genetic instruments by identifying any pleiotropic patterns. This framework was designed to reexamine and clarify the current MR assumptions to allow heterogeneous effect sizes and pervasive horizontal pleiotropy. A realistic assumption was made to say that horizontal pleiotropy is pervasive and that many variants may affect the outcome of the disease through their effects on other risk factors. The presence of pervasive horizontal pleiotropy can lead to a violation of the InSIDE (Instrument Strength Independent of Direct Effect) assumption which states that the strength of the instruments is independent of the direct effects on the outcome of interest and that the variant selected should only affect the outcome thought the risk factor of interest and no other pathway. The violation of this assumption is dealt with by more recent MR methods such as MRMix, Latent Causal Variable (LCV), Contamination mixture, and CAUSE, these methods allow a subset of instruments to be associated with a common pleiotropic pathway. These methods assume that for SNPs that violate the InSIDE assumption their pleiotropic effects would satisfy an equation. This more realistic assumption would allow for the determination of causal direction as well as the detection of numerous pleiotropic pathways. Thus, these methods can help account for multiple pleiotropic pathways to allow a more accurate analysis of the causal relationships between exposures and the outcome of interest.

GRAPPLE was used to analyse the effect of BMI, blood lipids, and systolic blood pressure on 25 different diseases. The GRAPPLE framework is an extension of a previous statistical framework called MR-RAPS (Mendelian Randomization Robust Adjusted Profile Score). This framework was shown to have many improvements over existing MR techniques, firstly, having the ability to efficiently use both strong and weak genetic instruments to determine causality. Weak instrument bias is avoided using profile likelihood to model pervasive pleiotropy if the InSIDE assumption is true for the majority of SNPs chosen. GRAPPLE also uses flexible p-value thresholds for instrument selection (as strict as $10^{-8}$ or as relaxed as $10^{-2}$) which helps avoid

bias and increases power[18]. Finally, this technique can detect numerous pleiotropic pathways and detect the causal direction of the effect as well as adjust for confounding risk factors through performing multivariable MR.

## MR in understanding the pathophysiology of Heart Failure and associated risk factors:

Heart failure has a multifactorial aetiology, meaning it is caused by the interactions of numerous environmental and genetic factors instead of a single factor. By understanding the causal relationships between risk factors/exposures such as biomarkers or individuals with therapeutic biomarkers and the development of heart failure within these people, we can gain further knowledge in the pathophysiology of heart failure, using Mendelian Randomisation. In this project, we will be focusing on applying Multivariable Mendelian Randomization to explore any direct causal links of the selected risk factors (LDL Cholesterol, coronary artery disease, Body Mass Index (BMI), Type 2 Diabetes and Systolic blood pressure) on the outcome Heart Failure, hopefully guiding the production of clinical interventions in future.

## MR Analysis Packages

Many packages have been made available to conduct Mendelian Randomization analysis, majority of them are available on R. A R package by MRCIEU (MRC Integrative Epidemiology Unit) called TwoSampleMR (version 0.6.3) allows you to perform MR analysis using GWAS summary data by using the IEU GWAS database. This package has a web app called MR-Base which allows you to try a limited range of this package's functionalities through their website. MendelianRandomisation (version 0.10.0) is another R package that was developed to carry out MR analyses by using summarised data and implements numerous methods to determine the causal effects of exposure on an outcome. The final R package is called MVMR (version 0.4) by WSpiller, this allows you to perform Multivariable MR analyses (including heterogeneity statistics for instrument strength and validity assessment).

## Methods

### Study Design:

Preliminary Univariable Two-Sample MR analyses will be conducted to act as a foundation before conducting Multivariable Two-Sample MR analyses. Univariable MR would allow the assessment of the causal relationship on the outcome for each exposure separately. Furthermore, it would allow the for the identification of pleiotropy and the violation of assumptions as well as help assess the strength of the genetic instruments being used. Subsequently, Multivariable MR would then be used to assess the joint effects of numerous exposures whilst accounting for confounding variables and mediation. Finally, robustness checks such as sensitivity analyses and horizontal pleiotropy tests would be used to increase confidence in any findings in this study. We will explore the causal relationships between coronary artery disease, LDL cholesterol, body mass index, systolic blood pressure, and type 2 diabetes mellitus on heart failure. The selected exposures and outcome data sources are available in the supplementary table below.

Univariable MR will be conducted with the R package TwoSampleMR, and Multivariable MR analyses will be conducted using the MVMR package. The MendelianRandomization package will be used to produce a robust estimate of the causal estimate for the IVW method in the Univariable MR analysis.

### Data Sources:

To conduct Two-Sample Univariable and Multivariable MR, GWAS summary statistics were collected for the outcome Heart Failure and each of the risk factors of interest. The statistics are publicly available resources on the GWAS Catalog. Summary statistics with a European ancestry were selected to prevent bias due to population stratification. Prior to data analysis, Liftover was performed using the package pyliftover (version 0.2) to ensure all genomic locations were aligned with the GRCh38 reference build to ensure reliable and valid results.

### Selecting Genetic Instruments and whether they violate core MR assumptions:

The GWAS summary data contains information such as the SNP rsID, beta coefficient (β), and p-values. The coefficient represents the effect size of the SNP on the trait, whilst the p-value represents the statistical significance of the association between the SNP and trait. SNPs with a p-value less than $5x10^{-8}$ were chosen to ensure the most statistically significant and robust genetic associations were considered within the study. This also helps mitigate the inclusion of false positive associations in the analysis. A lower threshold may help include stronger relevance but would introduce the susceptibility of weaker instruments.

The IV1 assumption (instrument strength), is the only assumption that can be explicitly tested using the F-Statistic. A mean cut-off of ≥10 would signify sufficient strength of the instrument[19]. The IV2 assumption (independence) uses a combination of tests such as weighted median and mode to assess pleiotropy and this assumption. Furthermore, MR egger can be used to estimate the directional pleiotropy and give a pleiotropy-robust estimate of the causal effect but should not be relied on as the only estimate as it's not as precise as the IVW method (When 100% of instruments are valid). As said previously, summary data chosen from the same population reduces the risk of population stratification, and minimising the overlap between the study populations can reduce bias of the effect estimate towards the observational estimate. The exclusion restriction (IV3) assumption can be tested using numerous methods, a common method is MR-Egger regression which allows you to check for directional horizontal pleiotropy based on the intercept of the regression line. A non-zero intercept would indicate the presence of horizontal pleiotropy.

**Harmonisation:**

Harmonisation is a critical step in Two-Sample MR analysis. This process involves aligning the effect alleles as well as scaling the effect estimates. Aligning the effect alleles ensures the effect of the variant on the exposure and the outcome are both measured in the same direction. For example, the effect allele for an SNP in the exposure data is 'G', and it is 'C' in the outcome dataset. The effect estimates and standard error for SNP in the outcome dataset need to be flipped. Scaling the estimates ensures the estimates are on the same scale, as some estimates could be in the units of standard deviations or millimetres. Thus, harmonisation allows the effect of the genetic variant to be measured correctly, in the correct direction, and allows for an accurate estimation (through scaling the effect estimates).

**LD Clumping and Palindromic SNPs:**

Linkage disequilibrium (LD) is the non-random association of alleles of different loci**.** These nearby genes can be inherited together and can all be associated with a certain trait. Due to LD, these SNPs will be tagging the same functional variant, and clumping these SNPs keeps the most significant SNP (based on p-value) to represent the group. We performed LD clumping with a $r^2$ threshold of 0.01 for Univariable MR and 0.001 for Multivariable MR, to ensure only independent genetic variants are retained for MR analysis**.**

Palindromic SNPs are a type of SNP where two possible alleles would complement each other on both the forward and reverse strands. These SNPs can provide a challenge when harmonising data in genetic studies. However, if allele frequencies of the SNP are available, it can be possible to infer the strand and the effect allele. The confidence of the inference is based on whether the allele frequency of the palindromic SNP is close to 0.5, if the frequency is close to 0.5 the effect allele may not be confidently inferred. If this is the case, or if allele frequency data is not available, excluding these SNPs from the analysis can be considered**.** For this analysis, we will try and infer the forward strand alleles using the allele frequency information.

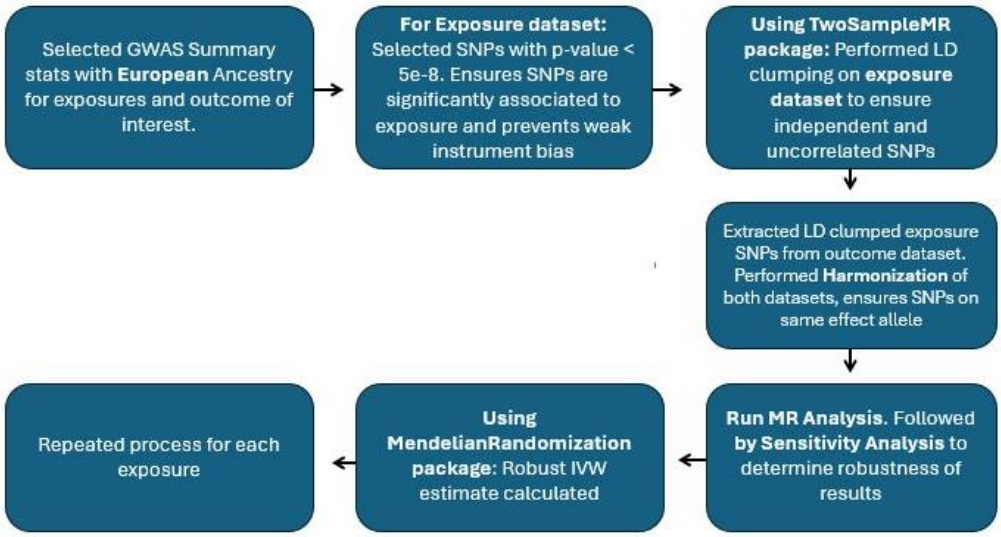



**Figure 4:** Flow chart displaying the process of how Two-Sample Univariable MR was conducted for each risk factor for the outcome Heart Failure.

**Estimation of pairwise covariances:**

Prior to conducting MVMR analysis, we had to estimate the pairwise covariances between the SNP associations. We estimated the covariance terms using the phenotypic correlation between the exposures. The "phenocov_mvmr()" function was used to estimate these terms. The function required two inputs: the phenotypic correlation between the exposures and the standard errors of the SNP-exposure effect sizes (betas).

## Results:

**SNP Selection:**

LD clumping was conducted for both Univariable and Multivariable MR, using $r^2$ of 0.01 and 0.001, respectively. Table 1 displays the number of SNPs after each stage of Univariable MR. The "Clumped SNPs" row indicates the number of SNPs remaining after performing LD clumping. The second row shows the number of SNPs remaining after harmonising the exposure data with the outcome data and filtering for MAF and any variants significantly associated with the outcome. During this step, we tried to infer the forward strand alleles using the allele frequency information for the palindromic SNPs. The final row is the number of SNPs that were used to estimate the total causal effect of the exposure on the outcome for the univariable analysis. For example, after filtering the LDL dataset for significant SNPs (p-value < $5x10^{-8}$) and performing LD Clumping, 714 significant and uncorrelated SNPs remained. These SNPs were harmonized with the outcome dataset and filtered with the parameters mentioned above to give 613 SNPs in the harmonized dataset. 603 of these SNPs were used to estimate the causal effect of LDL on Heart Failure.

| | BMI | LDL | CAD | SBP | T2DB |
|---|---|---|---|---|---|
| Clumped SNPs | 907 | 714 | 246 | 300 | 56 |
| SNPs after harmonisation | 900 | 613 | 224 | 291 | 56 |
| SNPs used | 877 | 603 | 216 | 288 | 56 |

**Table 1:** Displays the number of SNPs after clumping exposure data, harmonising exposure data with the outcome dataset, and the number of SNPs used in the Univariable Analysis.

**Univariable MR (UVMR) Analyses:**

The workflow for the Two-Sample Univariable analysis can be seen in Figure 4. After running MR analysis for all five exposures, significant associations were shown for at least two out of the four methods (MR-Egger, IVW, Weighted Median, and Simple Mode) for all exposures, except for Type 2 Diabetes, where the IVW method was

the only significant association. Our main method of interest was the Inverse-variance weighted method, and Table 2 displays the causal estimates for each exposure using this method. A p-value less than 0.05 would indicate a statistically significant association/relationship between the exposure and the outcome of Heart Failure. According to the UVMR results below, all these exposures display a significant causal relationship with Heart Failure using the IVW method. These results will be explained more thoroughly in the discussion section**.**

| Exposure | Outcome | Method | nsnp | Estimate(b) | se | pval | Lower CI | Upper CI |
|---|---|---|---|---|---|---|---|---|
| BMI | Heart Failure | IVW | 877 | 0.47 | 0.02 | 1.18E-98 | 0.43 | 0.51 |
| LDL Cholesterol | Heart Failure | IVW | 603 | 0.12 | 0.02 | 2.95E-11 | 0.09 | 0.16 |
| Coronary Artery Disease | Heart Failure | IVW | 216 | 0.32 | 0.02 | 1.75E-76 | 0.29 | 0.35 |
| Systolic Blood Pressure | Heart Failure | IVW | 288 | 0.33 | 0.05 | 7.54E-10 | 0.22 | 0.43 |
| Type 2 Diabetes Mellitus | Heart Failure | IVW | 56 | 0.08 | 0.02 | 1.00E-03 | 0.03 | 0.13 |

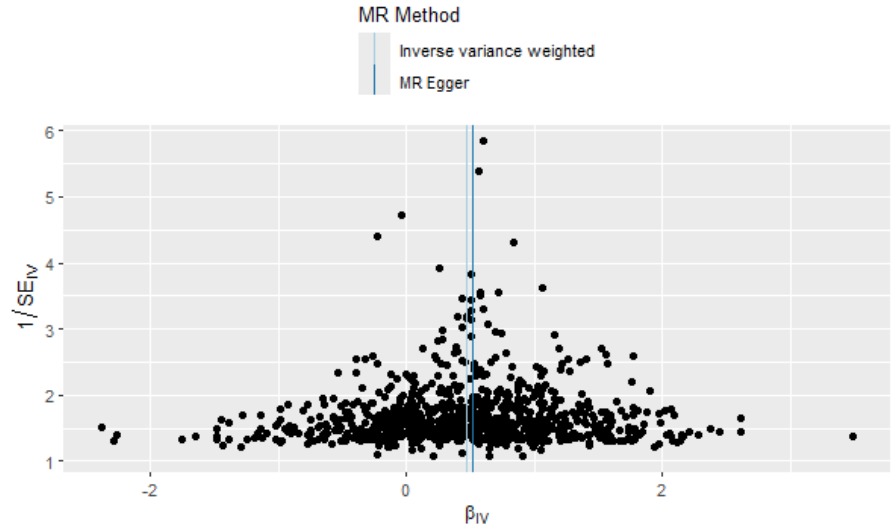
**Table 2:** MR results for the IVW estimates of the five exposures of interest on the outcome Heart Failure. The 95% Confidence intervals were calculated. Estimates represent the total causal effect of the selected exposure on Heart Failure.

**UVMR Sensitivity analyses:**

After initial Univariable analysis, results showed there to be significant associations between all the exposures and HF. However, for the IVW method, its causal estimate can be biased if pleiotropy is present in the analysis. Thus, we conducted sensitivity analyses to determine the robustness of our univariable results. The amount of pleiotropy in the analysis can be determined using the MR-Egger intercept (checks for exclusion restriction assumption). An intercept close to zero would indicate little pleiotropy present, and if the p-value is larger than 0.05, we can conclude there isn't any significant evidence of pleiotropy being present in the analysis. Table 3 displays the pleiotropy results for each exposure. For example, for BMI, there is shown to be a slightly negative MR-Egger intercept of –0.00076 with a standard error of 0.0011. The p-value of 0.50 (>0.05) indicates the pleiotropy present is not significant. A visual interpretation of the amount of pleiotropy present can be seen using a funnel plot. The funnel plot for the exposure BMI is displayed in Figure 5, an increased amount of asymmetry is directly proportional to the amount of pleiotropy present in that analysis. According to the results in Table 3, no significant pleiotropy is present amongst the instruments for each exposure.

| Exposure | Egger_Intercept | se | pval |
|---|---|---|---|
| BMI | -0.00076 | 0.0011 | 0.50 |
| LDL Cholesterol | -0.00054 | 0.00087 | 0.54 |
| Coronary Artery Disease | 0.0028 | 0.0022 | 0.21 |
| Systolic Blood Pressure | 0.0014 | 0.0021 | 0.52 |
| Type 2 Diabetes Mellitus | 0.0052 | 0.0042 | 0.22 |

**Table 3:** Pleiotropy results for the Univariable MR analyses for the exposures of interest. A p-value (pval) larger than 0.05 indicates no significant evidence of pleiotropy being present in the analysis.




f

**Figure 5:** Funnel plot for the exposure BMI. Asymmetry would indicate the presence of pleiotropy in the analysis.

The second stage of sensitivity analyses was checking for the presence of heterogeneity. Heterogeneity refers to the variability in the causal estimates across the genetic instruments used for the exposure. Heterogeneity can indicate the presence of invalid instruments and directional pleiotropy. Significant heterogeneity was shown for every exposure for the Univariable analyses. The original IVW causal estimate can be affected by heterogeneity and outliers, thus we calculated a robust IVW estimate using the MendelianRandomization R package to adjust for this using robust regression and penalized weights. Table 4 displays the adjusted IVW estimates for each exposure. Furthermore, the F-Statistics for all exposures are shown to be >10 indicating there is no weak instrumental bias present among the instruments used in the analysis when estimating the causal effect. The effect size estimates for both CAD and Systolic Blood pressure displayed an increase, from 0.32 and 0.33, to 0.33 and 0.35, respectively. However, the effect size of LDL cholesterol and Type 2 Diabetes Mellitus is shown to decrease from 0.12 and 0.08 to 0.11 and 0.07, respectively. These effect sizes were again shown to be significant due to the p-values being less than the 5% significance threshold.

| Exposure | Estimate(b) | se | Lower CI | Upper CI | pval | F-Stat |
|---|---|---|---|---|---|---|
| BMI | 0.47 | 0.02 | 0.43 | 0.51 | 0 | 53 |
| LDL Cholesterol | 0.11 | 0.02 | 0.07 | 0.16 | 0 | 136 |
| Coronary Artery Disease | 0.33 | 0.02 | 0.3 | 0.37 | 0 | 59 |
| Systolic Blood Pressure | 0.35 | 0.05 | 0.26 | 0.44 | 0 | 69 |
| Type 2 Diabetes Mellitus | 0.07 | 0.02 | 0.02 | 0.12 | 0 | 70 |

**Table 4:** Table displaying the Robust IVW estimates for each exposure. Used penalized weights and robust regression to account for outliers and heterogeneity.

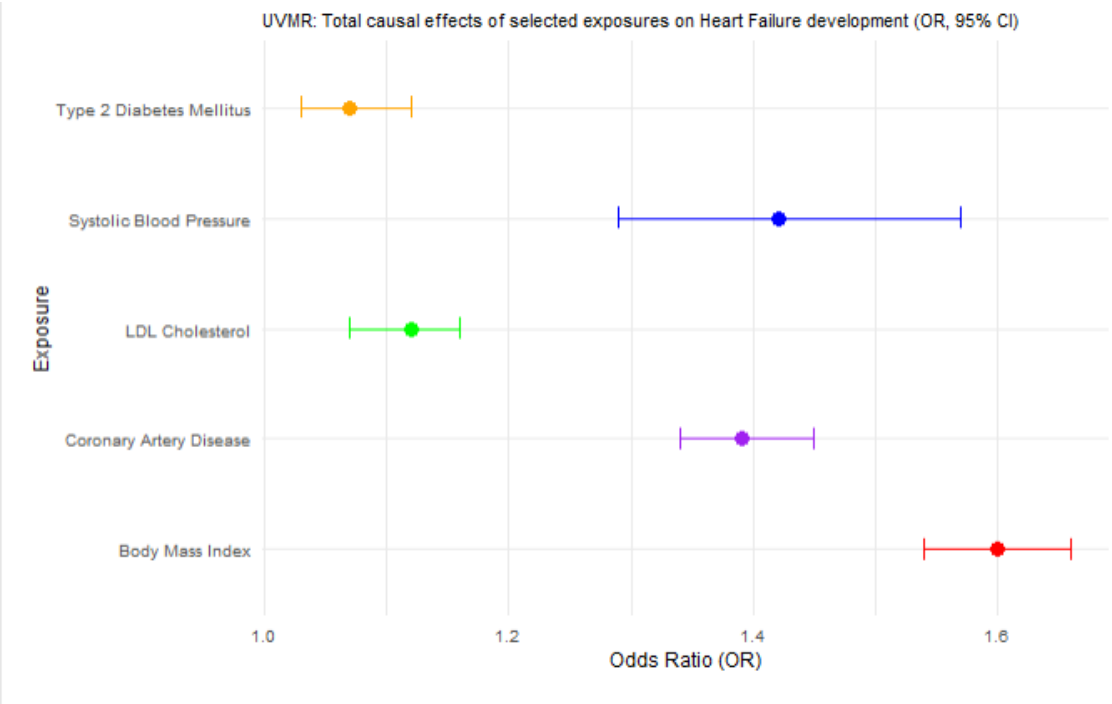



**Figure 6:** Displays odds ratio estimates and 95% CI for the total causal effect of each of the selected exposures on heart failure for the Univariable MR analysis. The robust IVW estimates from Table 4 were used to calculate the OR values and their confidence intervals.

**Multivariable MR:**

**MVMR Sensitivity Analyses:**

To determine the presence of weak instruments for MVMR analysis, we calculated the conditional F-Statistic for each exposure in our model. A statistic > 10 is the threshold to indicate sufficient strength of the instruments. Larger F-statistics would indicate the causal estimates are less likely to be biased (towards null hypothesis, of no causal effect) due to potentially weak instruments. The calculation was carried out using the strength_mvmr() function in the MVMR package. Table 5 displays the conditional F-Statistics for each exposure, they were all shown to be above 10 indicating no presence of weak instrument bias.

| | BMI | SBP | LDL | CAD | T2DB |
|---|---|---|---|---|---|
| F-Statistic | 11.81 | 56.20 | 39.20 | 17.60 | 16.30 |

**Table 5:** Conditional F-Statistic values for the selected exposures of interest. Pairwise covariance and MVMR formatted data are used as input for the strength_mvmr() function for the calculation. The same conventional instrument threshold of 10 used in the univariable analyses is used here.

After testing the strength of the instruments, we evaluated the presence of horizontal pleiotropy by testing for heterogeneity using Cochran's Q statistic. The pleiotropy_mvmr() function was used to calculate the potential heterogeneity and its significance. From our results, the high Q-statistic (245.06) and low p-value (<0.05) allowed us to conclude that there is significant heterogeneity present among the instruments. This in turn means some instruments included in the analysis may not be valid due to horizontal pleiotropy.

| Q-Statistic for Instrument validity | p-value |
|---|---|
| 245.06 on 143 DF | 2.32E-07 |

**Table 6:** Q-Statistic Estimate to evaluate the presence of heterogeneity. Observed heterogeneity can be indicative of a violation of the exclusion restriction assumption.

**Multivariable MR Analyses:**

In this MVMR model, the effect estimates are interpreted as the direct effect of each exposure on Heart Failure, conditional on the other exposures included in the model. From our results, BMI, Systolic blood pressure and coronary artery disease are shown to have statistically significant positive effects on Heart Failure. LDL cholesterol is shown to have a slight positive effect; however, this effect is not shown to be significant. Coronary artery disease is shown to have the most significant positive effect out of the five exposures included in the model. Furthermore, Type 2 Diabetes Mellitus is shown to have a very minimal positive effect, but this effect is shown to be not significant. However, these results are not robust to pleiotropic bias (due to heterogeneity being present in sensitivity analyses) and rely on other underlying MVMR assumptions having to be satisfied.

| Exposure | Estimate | Std.Error | t value | Pr(>\|t\|) | OR | L-CI OR | U-CI OR |
|---|---|---|---|---|---|---|---|
| Body Mass Index | 0.38 | 0.14 | 2.75 | 6.65E-03 | 1.46 | 1.11 | 1.92 |
| Systolic Blood Pressure | 0.14 | 0.07 | 2.01 | 4.63E-02 | 1.15 | 1.00 | 1.32 |
| LDL Cholesterol | 0.06 | 0.07 | 0.76 | 4.50E-01 | 1.06 | 0.93 | 1.22 |
| Coronary Artery Disease | 0.34 | 0.05 | 6.85 | 1.95E-10 | 1.40 | 1.27 | 1.55 |
| Type 2 Diabetes Mellitus | 0.04 | 0.03 | 1.22 | 2.23E-01 | 1.04 | 0.98 | 1.10 |

**Table 7:** Estimation of the direct causal effects for each exposure of interest. Calculations are done using the ivw_mvmr() function to use an inverse variance-weighted MVMR model to estimate direct causal effects. Odds ratios (OR) are calculated alongside the 95% Confidence intervals.

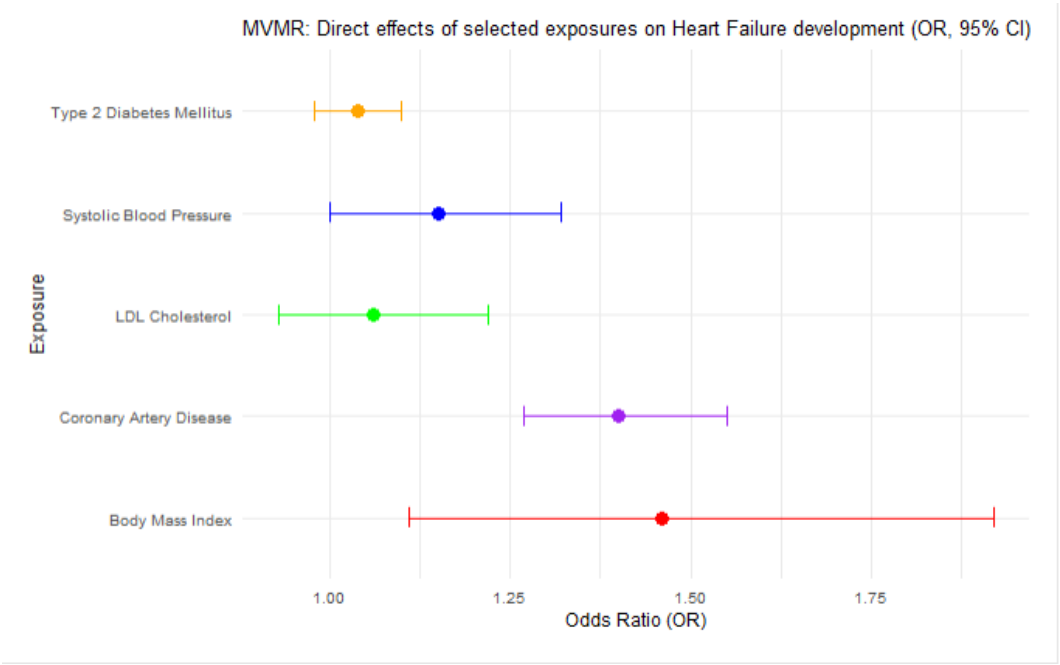



**Figure 7:** Displays the odd ratios for each exposure on the outcome of Heart Failure for initial MVMR analysis. Error bars indicate the 95% confidence intervals for each OR. Numerical values can be seen in Table 7.

**Robust causal effect estimation using Q-Statistic minimisation:**

The causal estimates we had calculated were shown not to be robust to pleiotropic bias due to significant heterogeneity being present in the analysis. When MVMR assumptions are potentially violated, we are able to calculate more robust causal estimates through Q-Statistic minimisation. These estimates can be seen in Table 8. The estimates for BMI and Diabetes remained approximately the same, whilst systolic blood pressure showed a significant decrease (0.10 in OR). Coronary artery disease also showed a 0.09 OR decrease. LDL Cholesterol showed a significant increase of 0.13 in OR (13% higher increase in developing HF). The 95% confidence intervals were not calculated due to them requiring a substantial amount of computing power.

| Exposure | Estimate | OR |
|---|---|---|
| Body Mass Index | 0.38 | 1.46 |
| Systolic Blood Pressure | 0.05 | 1.05 |
| LDL Cholesterol | 0.17 | 1.19 |
| Coronary Artery Disease | 0.27 | 1.31 |
| Type 2 Diabetes Mellitus | 0.07 | 1.07 |

**Table 8:** Displays robust estimates of the direct causal effects for the exposures of interest using Q-Statistic minimisation. Estimates calculated using the qhet_mvmr() function by minimising the Q-Statistic.

## Discussion:

This study aimed to determine if there were any significant causal relationships between the selected cardiovascular risk factors (BMI, Systolic Blood pressure, LDL cholesterol, coronary artery disease and Type 2 Diabetes Mellitus) and the development of Heart Failure using Multivariable Mendelian Randomization. Any findings in this experiment may illuminate possible therapeutic targets and enhance strategies in the prevention of heart failure. Preliminary univariable analysis was conducted as a foundation before conducting Multivariable MR. This was all followed by sensitivity analyses to determine the robustness of our results.

**Data selection reasoning:**

Summary statistics with a European ancestry were chosen to prevent bias due to population stratification. For Univariable MR, genetic variants with a p-value < $5x10^{-8}$ were selected to ensure the most significant variants were chosen for the analysis for each exposure. Furthermore, we filtered the harmonised data to remove any variants that have a "pval.outcome" value < $5x10^{-8}$, as they would violate the exclusion restriction assumption due to being significantly associated with the outcome. Finally, we excluded SNPs with minor allele frequencies (MAF) less than 5% as potential biases and errors can arise from these variants.

**Explaining OR values:**

The beta coefficients within the Heart Failure summary statistics were provided in the log-odds format. This is typically when a Logistic regression model is used. Therefore, effect sizes (b) from our results will be treated as log-odds values as in linear regression. To get a better understanding of this value, they can be converted into odds ratio (OR) values, and this can be done by exponentiating the log-odds value. OR values larger than 1 would indicate an increased risk of developing Heart Failure, whilst a value smaller than 1 would indicate a reduced risk. A value of 1 would mean the exposure has no overall effect on the development of Heart Failure.

**Univariable Results:**

Initial Univariable results indicated each exposure has a significant positive association with heart failure. However, sensitivity analyses showed significant heterogeneity to be present amongst the genetic instruments used to estimate the causal effect. Thus, we calculated robust IVW estimates using robust regression and penalized weights to adjust the estimates for heterogeneity and outliers.

The estimates calculated for the UVMR analysis represent the total causal effect of each exposure on the outcome. From Table 4, we can see the robust IVW effect estimate of LDL cholesterol on Heart Failure is 0.11 for the Univariable Analysis. Thus, by converting the log-odds value to odds ratio (1.12), we can interpret this as, a one-unit increase in LDL Cholesterol results in a 12% increase in the odds of developing Heart Failure. The data source for LDL Cholesterol measured LDL in millimoles per litre (mmol/L). Therefore, we can interpret this as a 1 mmol/L increase in LDL causes a 12% increase in the odds of developing Heart Failure. Figure 6 displays the odd ratios alongside the 95% Confidence intervals for each robust IVW estimate for the univariable MR analyses. BMI was shown to have the highest increase in the odds (60% increase in odds) of developing Heart Failure per one unit increase in the exposure, whilst Type 2 Diabetes Mellitus had the least (7% increase in odds per unit increase). Coronary artery disease showed a significant causal relationship with HF, with an OR value of 1.39 representing a 39% (95% CI, 34% - 45%) increase in the odds of developing heart failure. Finally, Systolic blood pressure was shown to cause a 42% (95% CI, 29% - 57%) increase in the odds of developing HF for a one-unit

increase in systolic blood pressure. The units for the one-unit increase for the remainder of the exposures still need to be determined in order to get a better understanding of the true causal effect.

**Multivariable Results:**

The causal effects for MVMR estimate the direct effects of each exposure on the outcome, conditional on the exposures included in the model. The effect estimates of LDL and Type 2 Diabetes Mellitus were shown to not be significant when accounting for the other exposures. The OR estimates for the initial MVMR results can be seen in Figure 7. Sensitivity analyses indicated the instruments were not prone to weak instrument bias, but pleiotropy may still be present. Robust estimates were calculated using Q-Statistic minimization; However, the confidence intervals were not calculated thus the OR values for the initial MVMR results were plotted.

The robust estimates showed reduced effect estimates for systolic blood pressure and coronary artery disease, while elevated values for Type 2 diabetes and LDL Cholesterol. Significant associations were seen between BMI, Systolic blood pressure, and coronary artery disease. In comparison to the UVMR results, both LDL Cholesterol and Type 2 diabetes causal relationship is no longer shown to be significant. This could be due to a reduction in the causal estimates from the UVMR analysis to MVMR, the reduction in the effect size could imply the original association (total effect size for UMR) was mediated or confounded by the exposures that are included in the MVMR model. However, further Mediation analysis would be required to determine whether the effects of one exposure are being mediated through another.

These results could allow the identification of potential targets to enhance the strategies to prevent heart failure. Exposures that are shown to have a significant effect on heart failure conditional to those included, could allow these exposures to be identified as potential treatment targets. By pinpointing these direct effects, MVMR would allow for more reliable targeted treatment as these exposures are shown to have an unmediated and direct effect on heart failure.

**Biological plausibility:**

Our results show significant associations between Body mass index, coronary artery disease, and systolic blood pressure with Heart failure. A study done in 2020 evaluated the causal effect of BMI on heart failure risk using Univariable and Multivariable MR analysis. Like our results, they concluded that BMI is significantly causally associated with an increased risk of HF (OR of 1.64 per 4.8kg/$m^2$ increase in BMI) for the Univariable MR analysis[20]. This was followed by MVMR analysis to estimate direct effects by accounting for potential mediators (atrial fibrillation, coronary artery disease, diabetes mellitus, and systolic blood pressure). The MVMR analysis revealed that the OR had dropped to 1.38 and concluded that 41% of the total effect of BMI was mediated by the exposures in the model. They concluded that a large proportion of the BMI-HF relationship is explained by BMI influencing these exposures in the model and that the residual direct effect may be explained by other pathways or exposures that weren't included in the model.

BMI has been shown to influence heart failure through numerous mechanisms such as Hemodynamic alterations (increased blood volume and cardiac output), hormonal effects of dysfunctional adipose tissue, and structural changes in the heart itself[21]. A higher BMI can also contribute to atherogenesis through the creation of a pro-inflammatory state and the excess circulating lipids can damage vascular tissues by lipotoxicity. These mechanisms may explain some of the mediated effects seen in this study and the one mentioned above.

One of the most significant variants used to estimate the total causal effect of BMI on heart failure was rs6567160. The nearest gene to this variant was melanocortin 4 receptor (MC4R), which is one of many that is located within the FTO region (obesity-associated). The most common cause of monogenic obesity is MC4R mutations which result in hyperphagia and weight gain. Agonists are currently in development to help regulate body weight by mimicking the natural function of MC4R, however, challenges have included the risk of side effects such as increased blood pressure.

Numerous studies have shown BMI to have a positive association with systolic blood pressure. Increased systolic blood pressure imposes a significant strain on the heart which in turn can lead to a series of events that result in heart failure. The increase in blood pressure causes an increased afterload, this is the amount of pressure the

heart needs to exert to eject blood out of the left ventricle[22] . To compensate for the increased workload, the left ventricular wall thickens (hypertrophy) to generate more force to pump the blood. Hypertrophy can lead to stiffness and weakened contractions over time which reduces the heart's ability to relax and fill with blood efficiently (diastolic dysfunction), as well as its ability to pump blood effectively (systolic dysfunction). Further mechanisms involving structural changes in the heart and activation of the renin-angiotensin-aldosterone system (RAAS) can lead to further straining and eventually heart failure[23]. These mechanisms can explain how a portion of the total BMI causal effect is mediated by systolic blood pressure.

LDL Cholesterol and Coronary artery disease together can play a role in the development of HF. When LDL is in excess within the blood it can penetrate the endothelium and enter the inside of the endothelium. Here LDL can become oxidized, and this oxidation process triggers an inflammatory response. This response results in monocytes being attracted to the area and being transformed into macrophages. These macrophages consume the oxidised LDL to become foam cells, these accumulate over time to become atherosclerotic plaques [24]. These plaques can grow, narrowing the arteries and reducing the blood flow toward the heart muscle, resulting in a condition called chronic ischemia which can damage the muscle. Rupturing of these atherosclerotic plaques can result in the complete blockage of a coronary artery which can result in myocardial infarctions. Repeated heart attacks and reduced blood flow can result in the progressive weakening of the heart muscle which eventually leads to heart failure.

Type 2 diabetes can lead to heart failure through many mechanisms including Endothelial dysfunction, which is where high sugar levels damage the endothelium, leading to atherosclerosis which is mentioned above. Other mechanisms involve damaged cardiac tissue due to diabetic cardiomyopathy which can result in inefficient pumping of blood around the body. In general, if not properly managed Type 2 diabetes can create a dangerous environment in which other conditions can arise.

We identified 56 significant SNPs that were used to estimate the total causal effect of Type 2 diabetes on Heart Failure. The most significant variant was rs7903146, which had a p-value of $1x10^{-200}$. Using Open Targets Genetics, we found its nearest gene to be TCF7L2 which is a transcription factor and downstream effector of the Wnt signalling pathway. This gene is shown to be strongly associated with type 2 diabetes due to being able to impair functions of pancreatic beta-cells (insulin secretion) and decrease the production of Glucagon-Like Peptide-1 (GLP-1) which affects how the body regulates its blood glucose levels. Small molecule inhibitors and RNA interference methods are being used to help modulate TCF7L2 activity by directly binding with the gene or preventing mRNA translation to reduce the production of the protein.

### Limitations of this Study:

This study used only lead variants and all available SNPs were not used. This may have resulted in important variants being excluded from the analysis. Furthermore, this study intended to use only European ancestry; However, the coronary artery disease summary statistics were shown to be predominantly European with additional data from a Japanese GWAS. As a result, bias would be present due to population stratification in this analysis.

### Future work:

This workflow used the lead variants when extracting the clumped exposure variants from the outcome dataset. Future work would be to incorporate the use of LD proxies to ensure a complete dataset. Using alternative variants with a high linkage disequilibrium would result in increased statistical power and robustness for any future analyses.

Furthermore, incorporating additional exposures to this MVMR model would allow for a more comprehensive understanding of the relationships and interactions between the exposures themselves, and the outcome of Heart Failure. Finally, individual-level data would provide a more accurate representation of the pairwise covariance terms used for MVMR. This would lead to increased precision and reliability when estimating the direct causal effects for each exposure.

Functional studies including gene expression studies and pathway analysis can be conducted on significant SNPs that have been identified for each exposure. Pathway analysis can help map these SNPs to associated genes to specific cardiovascular pathways to understand their biological significance. Another potential direction of future work is drug target Mendelian Randomization to explore the therapeutic potential of significant SNPs identified within this study. If an SNP is associated with the gene expression of a known drug target, it can be used to estimate the potential effect of modulating the target on the outcome. Thus, this would be effective for disease treatment and prevention by allowing the identification of genetic profiles for targeted therapies.

## Conclusion:

In conclusion, significant causal relationships were seen between body mass index, coronary artery disease, systolic blood pressure, and heart failure when estimating the direct causal effects. These exposures could allow for the identification of possible targets for future prevention treatments of heart failure due to having a direct and unmediated effect on the outcome. However, these results must factor in potential biases due to population stratification caused by coronary artery summary statistics.

## Acknowledgements:

I would like to express my gratitude to Dr Nay Aung and Hannah Nicholls for their insightful feedback and invaluable assistance with the MR analysis workflow.

## Supplementary Tables and Results:

The file attached contains information regarding the sources of the GWAS Summary Statistics used and the plots for each exposure for the univariable MR analysis. Supplementary Material.xlsx

## References:

1. Inamdar, A. and Inamdar, A. (2016). Heart failure: Diagnosis, management, and utilization. Journal of Clinical Medicine, 5(7), p.62. doi:https://doi.org/10.3390/jcm5070062.
2. Schwinger, R.H.G. (2021). Pathophysiology of heart failure. Cardiovascular Diagnosis and Therapy, 11(1), pp.263–276. doi:https://doi.org/10.21037/cdt-20-302.
3. National Health Service (2022). Heart failure. [online] NHS. Available at: https://www.nhs.uk/conditions/heart-failure/.
4. Shah, S., Henry, A., Roselli, C., Lin, H., Sveinbjörnsson, G., Fatemifar, G., Hedman, Å.K., Wilk, J.B., Morley, M.P., Chaffin, M.D., Helgadottir, A., Verweij, N., Dehghan, A., Almgren, P., Andersson, C., Aragam, K.G., Ärnlöv, J., Backman, J.D., Biggs, M.L. and Bloom, H.L. (2020). Genome-wide association and Mendelian randomization analysis provide insights into the pathogenesis of heart failure. Nature Communications, 11(1), p.163. doi:https://doi.org/10.1038/s41467-019-13690-5.
5. Bottle, A., Kim, D., Aylin, P., Cowie, M.R., Majeed, A. and Hayhoe, B. (2017). Routes to diagnosis of heart failure: observational study using linked data in England. Heart, 104(7), pp.600–605. doi:https://doi.org/10.1136/heartjnl-2017-312183.
6. Aung, N., Sanghvi, M.M., Piechnik, S.K., Neubauer, S., Munroe, P.B. and Petersen, S.E. (2020). The Effect of Blood Lipids on the Left Ventricle. Journal of the American College of Cardiology, 76(21), pp.2477–2488. doi:https://doi.org/10.1016/j.jacc.2020.09.583.
7. Hartley, A.E., Power, G.M., Sanderson, E. and Smith, G.D. (2022). A Guide for Understanding and Designing Mendelian Randomization Studies in the Musculoskeletal Field. JBMR Plus, 6(10). doi:https://doi.org/10.1002/jbm4.10675.

8. Davey Smith, G. and Hemani, G. (2014). Mendelian randomization: genetic anchors for causal inference in epidemiological studies. Human Molecular Genetics, 23(R1), pp.R89–R98. doi:https://doi.org/10.1093/hmg/ddu328.
9. Burgess, S. and Thompson, S.G. (2015). Multivariable Mendelian Randomization: The Use of Pleiotropic Genetic Variants to Estimate Causal Effects. American Journal of Epidemiology, 181(4), pp.251–260. doi:https://doi.org/10.1093/aje/kwu283.
10. Hemani, G., Bowden, J. and Davey Smith, G. (2018). Evaluating the potential role of pleiotropy in Mendelian randomization studies. *Human Molecular Genetics*, 27(R2), pp.R195–R208. doi:https://doi.org/10.1093/hmg/ddy163.
11. Louise, George Davey Smith and Tilling, K. (2024). Using the global randomization test as a Mendelian randomization falsification test for the exclusion restriction assumption. *European Journal of Epidemiology*. doi:https://doi.org/10.1007/s10654-024-01097-6.
12. Sanderson, E. (2020). Multivariable Mendelian Randomization and Mediation. Cold Spring Harbor Perspectives in Medicine, p.a038984. doi:https://doi.org/10.1101/cshperspect.a038984.
13. Sanderson, E., Spiller, W. and Bowden, J. (2021). Testing and correcting for weak and pleiotropic instruments in two-sample multivariable Mendelian randomization. Statistics in Medicine, 40(25), pp.5434–5452. doi:https://doi.org/10.1002/sim.9133.
14. mr-dictionary.mrcieu.ac.uk. (n.d.). *One-sample MR - Mendelian randomization dictionary*. [online] Available at: https://mr-dictionary.mrcieu.ac.uk/term/one-sample/.
15. mr-dictionary.mrcieu.ac.uk. (n.d.). Two-sample MR or MR with summary-level data - Mendelian randomization dictionary. [online] Available at: https://mr-dictionary.mrcieu.ac.uk/term/two-sample/ [Accessed 31 Jul. 2024].
16. Bowden, J., Davey Smith, G., Haycock, P.C. and Burgess, S. (2016). Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. Genetic Epidemiology, 40(4), pp.304–314. doi:https://doi.org/10.1002/gepi.21965.
17. Wootton, R.E., Lawn, R.B., Millard, L.A.C., Davies, N.M., Taylor, A.E., Munafò, M.R., Timpson, N.J., Davis, O.S.P., Smith, G.D. and Haworth, C.M.A. (2018). Evaluation of the causal effects between subjective wellbeing and cardiometabolic health: mendelian randomization study. BMJ, 362. doi:https://doi.org/10.1136/bmj.k3788.
18. Wang, J., Zhao, Q., Bowden, J., Hemani, G., Davey Smith, G., Small, D.S. and Zhang, N.R. (2021). Causal inference for heritable phenotypic risk factors using heterogeneous genetic instruments. PLOS Genetics, 17(6), p.e1009575. doi:https://doi.org/10.1371/journal.pgen.1009575.

19. Davies, N.M., Holmes, M.V. and Davey Smith, G. (2018). Reading Mendelian randomization studies: a guide, glossary, and checklist for clinicians. BMJ, p.k601. doi:https://doi.org/10.1136/bmj.k601.

20. R. Thomas Lumbers, Katsoulis, M., Henry, A., Mordi, I., Lang, C., Hemingway, H., Langenberg, C., Holmes, M.V. and Sattar, N. (2020). Body mass index and heart failure risk: a cohort study in 1.5 million individuals and Mendelian randomisation analysis. *medRxiv (Cold Spring Harbor Laboratory)*. doi:https://doi.org/10.1101/2020.09.23.20200360.

21. Ebong, I.A., Goff, D.C., Rodriguez, C.J., Chen, H. and Bertoni, A.G. (2014). Mechanisms of heart failure in obesity. *Obesity Research & Clinical Practice*, 8(6), pp.e540–e548. doi:https://doi.org/10.1016/j.orcp.2013.12.005.

22. LaCombe, P. and Lappin, S.L. (2020). *Physiology, Afterload Reduction*. [online] PubMed. Available at: https://www.ncbi.nlm.nih.gov/books/NBK493174/.

23. Orsborne, C., Chaggar, P.S., Shaw, S.M. and Williams, S.G. (2016). The renin-angiotensin-aldosterone system in heart failure for the non-specialist: the past, the present and the future. *Postgraduate Medical Journal*, 93(1095), pp.29–37. doi:https://doi.org/10.1136/postgradmedj-2016-134045.

24. Bentzon, J.F., Otsuka, F., Virmani, R. and Falk, E. (2014). Mechanisms of Plaque Formation and Rupture. *Circulation Research*, 114(12), pp.1852–1866.