# Celiac Disease Triggers

***Using Mendelian Randomization to infer causality***

The largest conducted Mendelian Randomization screening


**AUTHOR**
Matthijs Knigge

**INSTITUE**
Hanze University
Life Science and Technology

**ORGANIZATION**
Eriba
Deparment of Genetics


**DATE**
Wednesday, January
10-01-2018

# Celiac Disease Triggers

***Using Mendelian Randomization to infer causality***


**AUTHOR**
Matthijs Knigge

**SUPERVISORS**
Serena Sanna
Adriaan van der Graaf

**LECTURER**
Michiel Noback

**INSTITUE**
Hanze University
Life Science and Technology

**ORGANIZATION**
Eriba
Deparment of Genetics


**DATE**
Wednesday, January
10-01-2018

umcg

Hanze University of Applied Sciences Groningen

# PREFACE

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Aenean commodo ligula eget dolor. Aenean massa. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec quam felis, ultricies nec, pellentesque eu, pretium quis, sem. Nulla consequat massa quis enim. Donec pede justo, fringilla vel, aliquet nec, vulputate eget, arcu. In enim justo, rhoncus ut, imperdiet a, venenatis vitae, justo. Nullam dictum felis eu pede mollis pretium. Integer tincidunt. Cras dapibus. Vivamus elementum semper nisi. Aenean vulputate eleifend tellus. Aenean leo ligula, porttitor eu, consequat vitae, eleifend ac, enim. Aliquam lorem ante, dapibus in, viverra quis, feugiat a, tellus. Phasellus viverra nulla ut metus varius laoreet. Quisque rutrum. Aenean imperdiet. Etiam ultricies nisi vel augue. Curabitur ullamcorper ultricies nisi. Nam eget dui.

# ABSTRACT

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Aenean commodo ligula eget dolor. Aenean massa. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec quam felis, ultricies nec, pellentesque eu, pretium quis, sem. Nulla consequat massa quis enim. Donec pede justo, fringilla vel, aliquet nec, vulputate eget, arcu. In enim justo, rhoncus ut, imperdiet a, venenatis vitae, justo. Nullam dictum felis eu pede mollis pretium. Integer tincidunt. Cras dapibus. Vivamus elementum semper nisi. Aenean vulputate eleifend tellus. Aenean leo ligula, porttitor eu, consequat vitae, eleifend ac, enim. Aliquam lorem ante, dapibus in, viverra quis, feugiat a, tellus. Phasellus viverra nulla ut metus varius laoreet. Quisque rutrum. Aenean imperdiet. Etiam ultricies nisi vel augue. Curabitur ullamcorper ultricies nisi. Nam eget dui.

# ABBREVIATIONS

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Aenean commodo ligula eget dolor. Aenean massa. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec quam felis, ultricies nec, pellentesque eu, pretium quis, sem. Nulla consequat massa quis enim. Donec pede justo, fringilla vel, aliquet nec, vulputate eget, arcu. In enim justo, rhoncus ut, imperdiet a, venenatis vitae, justo. Nullam dictum felis eu pede mollis pretium. Integer tincidunt. Cras dapibus. Vivamus elementum semper nisi. Aenean vulputate eleifend tellus. Aenean leo ligula, porttitor eu, consequat vitae, eleifend ac, enim. Aliquam lorem ante, dapibus in, viverra quis, feugiat a, tellus. Phasellus viverra nulla ut metus varius laoreet. Quisque rutrum. Aenean imperdiet. Etiam ultricies nisi vel augue. Curabitur ullamcorper ultricies nisi. Nam eget dui.

# ORGANISATION

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Aenean commodo ligula eget dolor. Aenean massa. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec quam felis, ultricies nec, pellentesque eu, pretium quis, sem. Nulla consequat massa quis enim. Donec pede justo, fringilla vel, aliquet nec, vulputate eget, arcu. In enim justo, rhoncus ut, imperdiet a, venenatis vitae, justo. Nullam dictum felis eu pede mollis pretium. Integer tincidunt. Cras dapibus. Vivamus elementum semper nisi. Aenean vulputate eleifend tellus. Aenean leo ligula, porttitor eu, consequat vitae, eleifend ac, enim. Aliquam lorem ante, dapibus in, viverra quis, feugiat a, tellus. Phasellus viverra nulla ut metus varius laoreet. Quisque rutrum. Aenean imperdiet. Etiam ultricies nisi vel augue. Curabitur ullamcorper ultricies nisi. Nam eget dui.

# TABLE OF CONTENTS

# INTRODUCTION

Celiac disease is an autoimmune disorder that can have the possibility to occur genetically in predisposed people where the ingestion of gluten can cause damage in the small intestine. When people with celiac disease digest gluten their body triggers an immune response that attacks the small intestine. This causes damage to the villi that are lined up in the small intestine for nutrient uptake.[1] Celiac disorder affects up to two percent of the European population, and it is hereditary. People with a first-degree relative with celiac disease have the potential of developing celiac disease. The inheritance dogma of celiac disease is oligogenic, which means that a phenotypic outcome is determined by more than on gene. In this model, HLA-DQ2 is a genetic variant in the major histocompatibility complex (MHC) region that accounts for 40% of the inheritable risk. But, this genetic variant alone is not sufficient enough to develop celiac disease. 30% of the Europeans are carriers of the variant, however 3% express celiac disease.[2]

The aim of this study is the identification of factors that cause or protect against celiac disease, and to quantify the impact of the causal or protective relationship. To determine potential causal or protective relationships more than 440.000 public available clinical parameters and molecular mechanisms will be screened with the Mendelian Randomization (MR) approach. MR is an approach that uses genetic variants associated with a phenotype of interest to estimate a causal or protective relationship between this phenotype and a relevant medical outcome.[3] This study uses two-sample MR[4] methods; Inverse-Variance Weighted[5] (IVW) method and MR-Egger method[5].

We expect to find multiple causal or protective clinical parameters or molecular mechanisms that will be quantified to understand the impact of the causal or protective relationship, that can potentially explain what is triggering celiac disease in the 3% from the 30% of the Europeans that are carriers of HLA-DQ2.

# THEORY

## CELIAC DISEASE

Celiac disease is an autoimmune disorder that has the possibility to occur genetically in predisposed people where the ingestion of gluten cause damage. When people with celiac disease digest gluten their body triggers an immune response that attacks the small intestine.[1] This is characterized by small intestinal damage with the loss of function from the villi to take up nutrients, which leads to malabsorption.[6]

## MENDELIAN RANDOMIZATION

MR refers to the process of random segregation and assortment of alleles from an ancestor to offspring, that occur during gamete formation, in meiosis.[3, 4, 5]

This process gives us the advantage of using these genetic variants in observational settings to predict causal or protective relationships between exposure and outcome.[3, 4, 5] This process is comparable to Randomized Controlled Trials (RCT). In RCTs, the process of random and evenly distributing the sample into two arms ensures that known and unknown confounders are distributed evenly across both arms, see figure 1, stage (A). In MR, the random segregation and assortment of alleles is analogous the process of randomization in RCTs. In stage (B) for RCT one arm will be the case group, where some form of experiment will be conducted, and the other is the control group. At this stage for MR, both arms will represent different genotypes which in stage (C) will lead to different products. For RTC at this stage, in the case arm there will take place on-off-target effects. At the final stage (D), differences between the two arms from both RCT, and MR can be studied and should only reflect the differences between the two arms.

MR uses genetic variant for a variety of reasons. (I) genetics variants are less amenable for confounding factors. As denoted by Mendel's first law, the law of segregation; genetic variants segregate randomly and independently from environmental factors, and Mendel's second law, the law of independent assortment; genetic variants segregate independently from other traits. (II) reverse causality does not affect the genetic instruments because an individuals germline genotype precedes the medically relevant disease.[3, 4, 5] (III) genetic variants are subjected to little measurement error or bias. (IV) MR does not force to use the actual causal variant but is satisfied with a marker that is in Linkage Disequilibrium with the causal genetic variant. (V) MR gives us the advantage to use the genetic variants from Genome-Wide Association Study (GWAS) summary statistics, which are routinely available on large well-phenotyped scale.[7, 8]

The MR approach uses genetic variants as Instrumental Variables (IVs) to infer a causal or protective relationship between an exposure and outcome. IVs are variables that are associated with the exposure phenotype of interest and does not suffer from reverse causality or confounding factors.[3, 4, 5] Reverse causality is the phenomena where it is expected that for example X causes changes in Y, but it can be that Y causes a change in X.

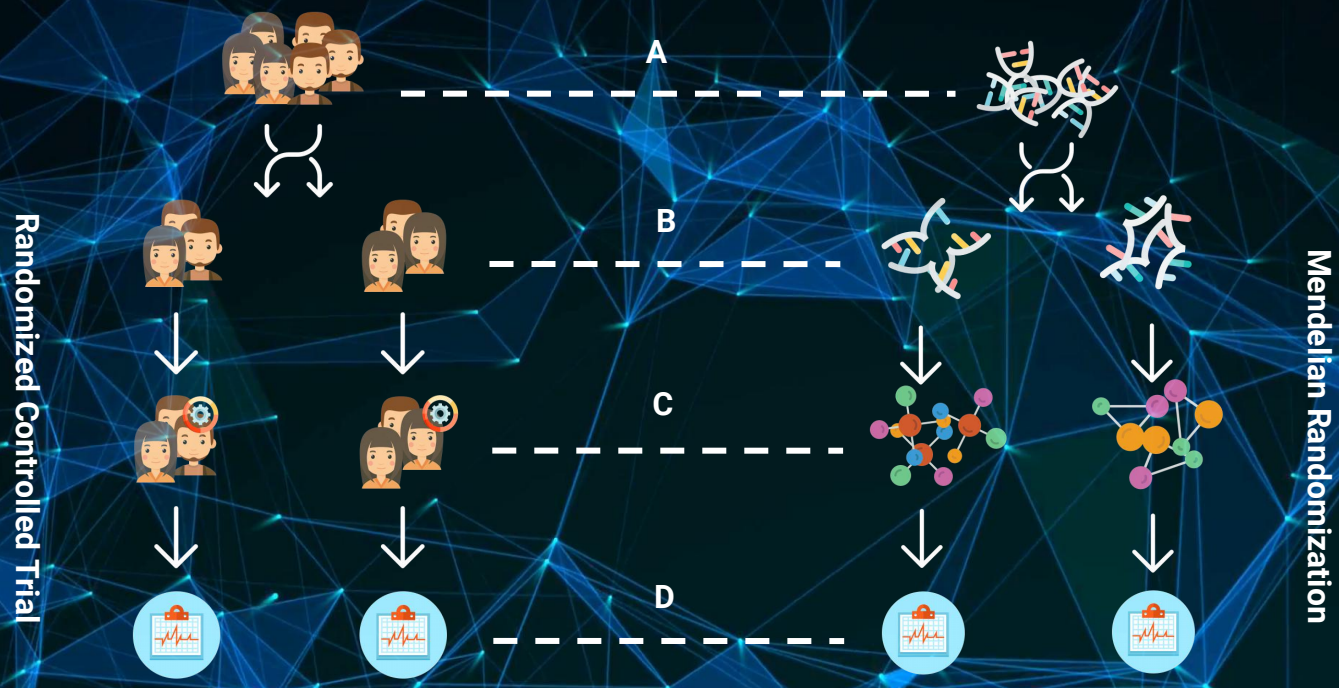


**Figure 1: RCT in comparison to MR. The left diagram is an overview of a RCT, and the right diagram is an overview of a MR. (A) describes the random allocates and segregation of alleles in MR, and in RCT the random distribution of the population into two arms. At stage (B); here in RCT there will be a case group and a control group, and for MR there will be two different genotypes. After that, stage (C); for RCT here will take place different target effects for one of the arms, for MR there will be two different products from the genotype which in turn in stage (D) give rise to a different effect, as does this take place for RCT. And these differences in arms can be studied.**

Take this example: an individual who characterized by being a heavy alcohol consumer, and because of the over consumption is that individual depressed. But, it is possible that this individual is heavy alcohol consumer due to being depressed. Confounding is the variable that is not accounted for and can bias the analysis by suggesting that there is correlation when that is not the case.

For a genetic variable to be used as an IV, the variant must meet three core assumptions. (I) the genetic variant must be associated with the exposure, (II) the genetic variant must not be associated with confounders, (III) the genetic variable can only be associated with the outcome through the exposure[3, 4, 5], see figure 2.

The first assumption can be assessed by investigation of the strength of the association between the genetic instrument and the exposure. The second assumption can be assessed by examining the relationship between the genetic variant and potential confounders. Final, the third assumption can be addressed by examining the biological pathways of the genetic instruments.[3]

## ONE-SAMPLE MENDELIAN RANDOMIZATION

One-sample MR is a method for predicting the causal effect between outcome and exposure. This method restricts the analysis to draw both, outcome and exposure from the same population sample. To calculate the causal estimate between the outcome and exposure one needs to take the ratio of the effect from

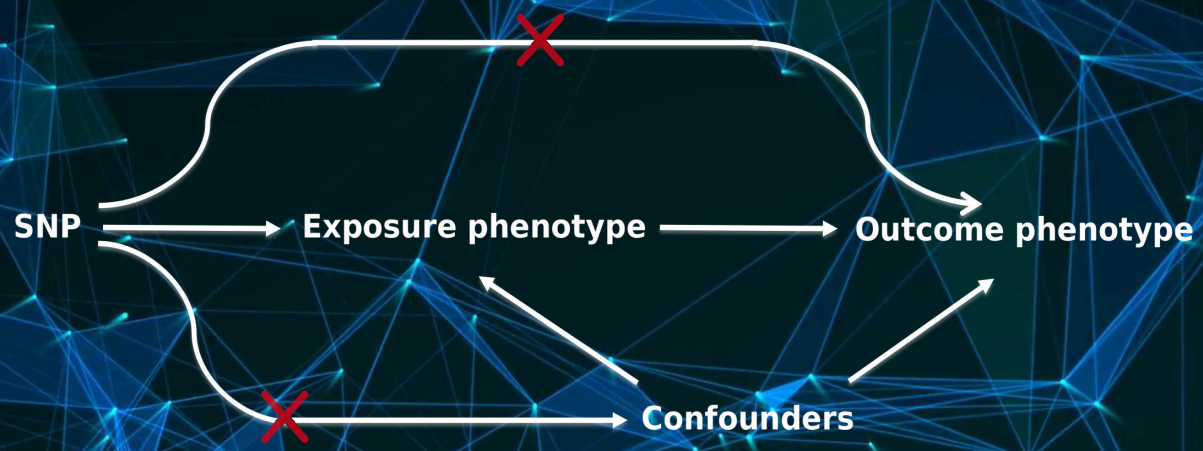


**Figure 2: graph showing the assumptions made by MR. The nodes in this graph represent the genetic instrument: Single-nucleotide polymorphism (SNP), the exposure, the outcome, and confounders. This graphs shows the relationships between variables that meet the three core assumptions. (I) Genetic variant must be associated with exposure, (II) no association between genetic variant and confounders, (III) only association with outcome through exposure.**

the genetic variant on outcome divided by the effect from the genetic variant on the exposure.[4] This is referred to as the Wald method ($\hat{\beta}_{IV}$), see equation 1. Where $\hat{\beta}_{Y|G}$ is the effect from the genetic variant on outcome, and $\hat{\beta}_{X|G}$ is the effect from the genetic variant on the exposure.[3, 4]

$$\hat{\beta}_{IV} = \frac{\hat{\beta}_{Y|G}}{\hat{\beta}_{X|G}} \qquad (1)$$

## TWO-SAMPLE MENDELIAN RANDOMIZATION

Two-sample MR takes advantage of the fact that it is not needed to obtain both, the exposure and outcome from the same population sample. The two-sample approach method makes it possible to use publicly available GWAS.

# MATERIAL & METHODS

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Aenean commodo ligula eget dolor. Aenean massa. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec quam felis, ultricies nec, pellentesque eu, pretium quis, sem. Nulla consequat massa quis enim. Donec pede justo, fringilla vel, aliquet nec, vulputate eget, arcu. In enim justo, rhoncus ut, imperdiet a, venenatis vitae, justo. Nullam dictum felis eu pede mollis pretium. Integer tincidunt. Cras dapibus. Vivamus elementum semper nisi. Aenean vulputate eleifend tellus. Aenean leo ligula, porttitor eu, consequat vitae, eleifend ac, enim. Aliquam lorem ante, dapibus in, viverra quis, feugiat a, tellus. Phasellus viverra nulla ut metus varius laoreet. Quisque rutrum. Aenean imperdiet. Etiam ultricies nisi vel augue. Curabitur ullamcorper ultricies nisi. Nam eget dui.

.

# RESULTS

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Aenean commodo ligula eget dolor. Aenean massa. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec quam felis, ultricies nec, pellentesque eu, pretium quis, sem. Nulla consequat massa quis enim. Donec pede justo, fringilla vel, aliquet nec, vulputate eget, arcu. In enim justo, rhoncus ut, imperdiet a, venenatis vitae, justo. Nullam dictum felis eu pede mollis pretium. Integer tincidunt. Cras dapibus. Vivamus e

# DISCUSSION

# CONSLUSIONS

# REFERENCES


1. Robert Di Niro, et al. High abundance of plasma cells secreting transglutaminase 2-specific IgA autoantibodies with limited somatic hypermutation in celiac disease intestinal lesions. Nature Medicine 18, 441-445 (2012). https://doi.org/10.1016/S0016-5085(98)70008-3

2. Gujral, N., Freeman, H. J., & Thomson, A. B. (2012). Celiac disease: Prevalence, diagnosis, pathogenesis and treatment. World Journal of Gastroenterology : WJG, 18(42), 6036–6059. http://doi.org/10.3748/wjg.v18.i42.6036

3. David M. Evans and George Davey Smith. Mendelian Randomization: New Applications in the Coming Age of Hypothesis-Free Causality. Annual Review of Genomics and Human Genetics 16, 327-350. http://www.annualreviews.org/doi/10.1146/annurev-genom-090314-050016

4. Debbie A. Lawlor. Commentary: Two-sample Mendelian randomization: opportunities and challenges. International Journal of Epidemiology, Volume 45, Issue 3, 1 June 2016, Pages 908–915, https://doi.org/10.1093/ije/dyw127

5. Stephan Burgess, Simon G. Thompson. Interpreting findings from Mendelian randomization using the MR-Egger method. S.G. Eur J Epidemiol (2017) 32: 377. https://doi.org/10.1007/s10654-017-0255-x

6. Jerry S, Trier, M.D. Celiac Sprue. N Engl J Med 1991; 325:1709-1719December 12, 1991 http://www.nejm.org/doi/full/10.1056/NEJM199112123252406

7. George Davey Smith, Gibran Hemani. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. Human Molecular Genetics, Volume 23, Issue R1, 15 September 2014, Pages R89–R98, https://doi.org/10.1093/hmg/ddu328

8. Caroline L Relton, George Davey Smith. Two-step epigenetic Mendelian randomization: a strategy for establishing the causal role of epigenetic processes in pathways to disease. International Journal of Epidemiology, Volume 41, Issue 1, 1 February 2012, Pages 161–176, https://doi.org/10.1093/ije/dyr233

# APPENDIX

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Aenean commodo ligula eget dolor. Aenean massa. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec quam felis, ultricies nec, pellentesque eu, pretium quis, sem. Nulla consequat massa quis enim. Donec pede justo, fringilla vel, aliquet nec, vulputate eget, arcu. In enim justo, rhoncus ut, imperdiet a, venenatis vitae, justo. Nullam dictum felis eu pede mollis pretium. Integer tincidunt. Cras dapibus. Vivamus elementum semper nisi. Aenean vulputate eleifend tellus. Aenean leo ligula, porttitor eu, consequat vitae, eleifend ac, enim. Aliquam lorem ante, dapibus in, viverra quis, feugiat a, tellus. Phasellus viverra nulla ut metus varius laoreet. Quisque rutrum. Aenean imperdiet. Etiam ultricies nisi vel augue. Curabitur ullamcorper ultricies nisi. Nam eget dui.Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Aenean commodo ligula eget dolor. Aenean massa. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec quam felis, ultricies nec, pellentesque eu, pretium quis, sem. Nulla consequat massa quis enim. Donec pede justo, fringilla vel, aliquet nec, vulputate eget, arcu. In enim justo, rhoncus ut, imperdiet a.