UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE ESTUDIOS ESTADÍSTICOS

TESIS DOCTORAL

Statistical techniques for estimating causal effects in biomedical research

Técnicas estadísticas para la estimación del efecto causal en investigación biomédica

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Claudia Coscia Requena

Directoras

Teresa Pérez Pérez
Núria Malats Riera

Madrid

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE ESTUDIOS ESTADÍSTICOS

TESIS DOCTORAL

Statistical techniques for estimating causal effects in biomedical research

Técnicas estadísticas para la estimación del efecto causal en investigación biomédica

MEMORIA PARA OPTAR AL GRADO DE DOCTORA

PRESENTADA POR

Claudia Coscia Requena

DIRECTORAS

Teresa Pérez Pérez

Núria Malats Riera

Madrid, Diciembre 2021

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE ESTUDIOS ESTADÍSTICOS

TESIS DOCTORAL

Statistical techniques for estimating causal effects in biomedical research

Técnicas estadísticas para la estimación del efecto causal en investigación biomédica

MEMORIA PARA OPTAR AL GRADO DE DOCTORA

PRESENTADA POR

Claudia Coscia Requena

DIRECTORAS

Teresa Pérez Pérez
Núria Malats Riera

Madrid, 2021

A mis abuelas, Fuensanta y Filomena

A mis padres

A Matías

Agradecimientos

Después de estos cuatro años de tesis, no podría empezar de otra manera sino agradeciendo a mis directoras de tesis, Teresa y Núria. Teresa, gracias por el apoyo que me has dado desde los comienzos de la tesis. Núria, gracias por haberme dado la oportunidad de formar parte de tu grupo y por haberme apoyado a nivel personal y profesional. A ambas, gracias por vuestra confianza y por haberme enseñado tanto.

También me gustaría agradecer a los que fueron mis directores de tesis, Alfonso y Esther. Alfonso, fuiste quién me animó a empezar esta carrera científica. Gracias por introducirme en el mundo de la inferencia causal. Esther, aunque breve, en nuestra etapa como directora-doctoranda he aprendido muchísimo, gracias por todo el apoyo que me diste y me sigues dando. ¡Tenemos un "*Nature*" pendiente!

To Stephen Burgess, thank you for hosting me during 2021. It has been an incredible experience for me, and I learned a lot from you and your group. I hope to finally meet you in person one day!

A todo el grupo de Epidemiología Genética y Molecular del CNIO. ¡Gracias por tanto! No tengo palabras para expresar lo que ha significado para mí trabajar a vuestro lado, sois un grupo extraordinario. Evangelina, gracias por todos tus consejos, y por haberme ayudado siempre que lo he necesitado. Alberto, Fran, Harry y Sergio, gracias por vuestras risas y buen rollo durante estos años, y por los buenos ánimos que siempre me habéis dado! Vanesa, ha sido un placer trabajar contigo y tener nuestras charlas en los descansos de trabajo. ¡Lidia, qué maravilla haber coincidido contigo y haberte conocido! Gracias por tu cariño, por tus abrazos en los momentos más difíciles. Echaré mucho de menos el momento diario de "¿vienes a por agua?". Espero seguir conectadas aún después de la tesis. Lola, para mi has sido más que una compañera de trabajo, desde el primer momento me hiciste sentir en casa. Echaré muchísimo de menos nuestros FI, las turrécdotas, los paseíllos yendo a Chamartín y nuestros audios interminables (¡aunque esto espero que lo sigamos haciendo!). Grazie per tutto, amica mia! Raquel, mi casi-melliza en esta etapa predoctoral. No cabe en este agradecimiento lo que has significado para mí, dentro y fuera del CNIO. Has sido mi apoyo constante desde el principio, siempre animándome a ser mejor investigadora y apoyándome frente a cualquier obstáculo que he tenido, aun con tus risas por mis gomas de borrar con forma de corazón, la manera de ponerme la bufanda y sentarme en la silla… Echaré mucho de menos no tenerte como compi de banco, y que con una mirada cómplice nos entendamos a la primera, y, sobre todo, echaré de menos nuestras risas. ¡Te espero en Lyon para comer queso *president* con boina francesa incluida! Gracias también a todos los miembros antiguos y nuevos de GMEG, Olga, Miguel, Barbara, Dina, Nannan, Ashwag… ¡Todos a vuestra manera habéis hecho que esta experiencia haya merecido la pena!

Como no, mil gracias a las amigas que me llevo del CNIO. Mago, Jessi, Thelma, Mónica. Habéis sido un pilar fundamental durante esta etapa, gracias por las risas, por los cafés en la social room, por los llantos compartidos y por todos los buenos momentos pasado juntas. Maguito, por tu buen rollo constante y por ayudarme a ver las cosas siempre desde otra perspectiva. Jessi, hufflepuff aunque lo niegues, gracias escucharme siempre y por haber estado en los momentos buenos, y no tan buenos. Moniquet, por haberme transmitido tu optimismo y buen humor siempre. Thelmi, por las risas juntas, echaré mucho de menos no mandarnos corazones desde la otra

punta de la cafetería. No tengo palabras para deciros todo lo que os quiero, y lo maravilloso que ha sido que entrarais en mi vida. ¡Por más viajes, excursiones y posteriores cenas mexicanas! ¡Y por supuesto, por más 8M juntas!

A las amigas que estaban desde antes del CNIO. Adri, gracias por haber estado durante todos estos años, por escuchar conmigo a Rino Gaetano y Vecchioni, por nuestras Nocheviejas en Madrid y por todas las conversaciones que me han hecho crecer tanto. Eli, empecé esta tesis cuando tú empezabas el grado, y tengo la sensación de haber crecido a tu lado en este recorrido. Ici, pequeña padawan, gracias por las carcajadas, por los bailes y por el buen rollo que siempre me has transmitido. Gracias por todas las charlas, cervezas y fiestas juntas. ¡Habéis hecho de esta etapa una experiencia mucho más amena!

A mi familia española. Gracias a mis tíos por el apoyo que me han dado estos diez años en Madrid. Miguel, por haber sido mi mentor académico, por haberme echado una mano y haberme dado consejos siempre que los he necesitado. Abuelo, gracias por ser mi fan número uno desde que tengo uso de razón y por haber creído en mí siempre. Gracias por leerme a Machado, por hablarme de Granada y de Murcia con la abuela, por cantarme coplas y contarme tus anécdotas de toda una vida. Gracias por todos los buenos momentos vividos a tu lado.

A mi familia italiana. Grazie, Marilù, per esserci sempre anche se nella distanza, per essere stata come una sorella e amica confidente in tutti questi anni.

A toda mi familia adquirida en Uruguay. En especial a mis suegros, Adriana y Pablo, por haberme hecho sentir parte de la familia desde el principio, por habernos apoyado siempre a Matías y a mí en nuestro recorrido.

A mis hermanos, Michele y Antonio. Pese al desafío de la convivencia juntos (¡no se puede negar!), no me imagino este recorrido sin vosotros a mi lado. Miki, me acompañaste ya en mi defensa del TFG y TFM, creyendo en mí y estando siempre en los momentos más importantes. Anto, me has enseñado que, con esfuerzo y fuerza de voluntad, todo se puede conseguir. A ambos os debo mucho, gracias por todo.

A mi madre, Rocío, por haberme enseñado desde pequeña a creer en mí misma y en hacerme valer. Has sido y eres uno de los pilares más importantes en mi vida, y un modelo a seguir como mujer y madre. Gracias por haberme enseñando a ser perseverante y decidida, y que un poco de paciencia y optimismo, todo tiene solución.

A mio padre, Gaetano, e a Saveria, per tutto quello che avete fatto per noi sempre, per avermi dato il supporto di cui avevo bisogno. Papà, grazie per avermi trasmesso tanti buoni valori e il senso della famiglia, per avermi insegnato che, con sacrificio ed entusiasmo, si riesce a far tutto. Una volta mi dicesti: "ricorda, anche se sei piccoletta, hai cento metri di radici che ti tengono". Oggi, non ho dubbi che siete stati tutti voi a reggermi nei momenti piú duri.

Por último, no menos importante, a Matías. Gracias por haber priorizado siempre este proyecto, por todos tus "primero la tesis y luego vemos", por haber sido el mejor compañero y apoyo incondicional durante estos cuatro años. Como dice la canción, "me prestaste un beso, me prestaste calma, me prestaste todo, lo que me faltaba". ¡Que empiece nuestra próxima aventura!

Index

Index of Tables, Figures, and Boxes

Index of Main Tables

Index of Main Figures:

Index of Boxes

Glossary and abbreviations

RCT: Randomized Controlled Trial

X: Treatment, risk factor or exposure

Y: Outcome

Y(x): Potential Outcome

ICE: Individual causal effect

ACE: Average causal effect

Z: Set of covariates

MCRD: Marginal Causal Risk Difference

DAG: Directed Acyclic Graph

C: Set of confounders or colliders

CsI: Causal Inference

CMA: Causal Mediation Analysis

MR: Mendelian Randomization

PS: Propensity Score

IPTW: Inverse Probability Treatment Weighting

W: Weights for the Propensity Score

M: Mediator

NDE: Natural Direct Effect

NIE: Natural Indirect Effect

TE: Total Effect

Y(x,M(x)): Potential outcome according to X and M

IV: Instrumental variable

X(IV): Potential outcome for X according to IV

SNP: Single Nucleotide Polymorphism

G: Single SNP or allele score

2SLS: Two-stage-least-square

LD: Linkage Disequilibrium

IVW: Inverse Variance Weighted

MVMR: Multivariable Mendelian Randomization

ICU: Intensive Care Units

VTE: Venous Thromboembolism

PE: Pulmonary embolism

DVT: Deep Vein Thromboembolism

PC: Pancreatic ductal adenocarcinoma

T2DM: Type 2 diabetes mellitus

NODM: New onset diabetes mellitus

LSDM: Long standing diabetes mellitus

BMI: Body mass index

BLCA: Bladder cancer

MRinCMA: Mendelian Randomization in Causal Mediation Analysis

4SLS: Four-stage-least-squares

4SRI: Four-stage-residual-inclusion

3SLS: Three-stage-least-square

3SRI: Three-stage-residual-inclusion

SEM: Structural Equation Model

MSE: Mean Square Error

EmpSE: Empirical Standard Error

CI: Confidence Interval

CI-C: Coverage

BE-C: Bias-eliminated Coverage

Abstract / Resumen

Abstract

Causal inference methods are statistical techniques used to analyse the causal effect of a treatment/exposure on an outcome. Their use is increasing in the last decade, especially in the framework of observational studies where the no randomization of the treatment/exposure may lead to confounding bias. These methods present great advantages versus classic regression models due to their capability of reducing and controlling for confounding bias.

This thesis begins with the use of known techniques applied in real clinical scenarios, second, a lack of developed statistical methods to estimate causal effects in complex epidemiological scenarios is noted. These findings support the main objective of this thesis, which is the development of causal inference methods to better understand and diagnose clinical and epidemiological outcomes.

A comparison between the Propensity Score and classic regression models was made using an Intensive Care Unit database where it was shown that, in presence of confounding bias, Propensity Score performed better. Moreover, based on a systematic review and meta-analysis, causal estimates from Propensity Score and Randomized Controlled Trials were compared. It was observed that similar estimations were obtained in both approaches.

The Mendelian Randomization and Causal Mediation Analysis techniques were applied to study the causal effect of type 2 diabetes mellitus subtypes (new-onset, and long-standing, respectively) and obesity on pancreatic cancer. While long-standing diabetes did not present evidence of causality on pancreatic cancer, pancreatic cancer showed a causal effect on new-onset diabetes. Moreover, obesity acted as a mediator for the effect of type 2 diabetes mellitus on pancreatic cancer. These results justified the need for the development of new statistical

techniques to further study the complex relationship between diabetes, obesity, and pancreatic cancer.

First, a new statistical methodology, *MRinCMA*, was proposed. This is an extension of Causal Mediation Analysis using Mendelian Randomization, to obtain unbiased estimates in the mediation framework. Using simulation studies, it was shown that *MRinCMA* provided unbiased estimates in comparison to the Structural Equation Models, used as the reference method. This approach was applied to study the effect of obesity on pancreatic cancer considering long-standing diabetes as a mediator where no evidence for causality was found.

Second, an extension of Mendelian Randomization which allows researchers to stratify the population without introducing collider bias was proposed. Using simulation studies, we proved that the new approach was able to obtain unbiased estimates and to detect heterogeneity across strata of the population.

Overall, we have shown the importance of correctly applying causal inference approaches and their differences with classic regression models. Moreover, we have proposed some extensions to the classic Mendelian Randomization that can be used in other biomedical scenarios.

Resumen

Los métodos de inferencia causal son técnicas estadísticas utilizadas para analizar el efecto causal de un tratamiento/exposición sobre un desenlace de interés. Su uso ha aumentado en la última década, especialmente en estudios observacionales donde la no aleatorización del tratamiento deriva en resultados sesgados. Esta metodología presenta enormes ventajas frente a modelos de regresión clásicos, debido a su capacidad de reducir y ajustar por el sesgo de confusión.

Esta tesis comienza con la aplicación de técnicas ya conocidas en escenarios clínicos donde se muestra la falta de desarrollo estadístico para estimar el efecto causal en escenarios biomédicos más complejos. Estos hallazgos justifican el objetivo principal de esta tesis que es desarrollar métodos de inferencia causal para entender y diagnosticar mejor eventos clínicos y epidemiológicos.

Se comparó el Índice de Propensión con los modelos clásicos de regresión, utilizando una base de datos de pacientes ingresados en UCI y se observó que, en presencia del sesgo de confusión, el Índice de Propensión aportó mejores resultados. Además, mediante una revisión sistemática y un meta-análisis, se compararon las estimaciones causales obtenidas aplicando el Índice de Propensión en estudios observacionales y las obtenidas en Ensayos Clínicos Aleatorizados. Se observó que en ambos casos las estimaciones obtenidas eran muy similares.

Las técnicas de *Aleatorización Mendeliana* y Análisis de Mediación Causal se aplicaron para estudiar el efecto causal de los subtipos de la diabetes mellitus tipo 2 (temprana y tardía, respectivamente) y de la obesidad sobre el cáncer de páncreas. Se concluyó que, mientras la diabetes tardía no presentaba un efecto causal sobre el cáncer de páncreas, éste último sí que

tenía un efecto sobre la diabetes temprana. Además, se observó que tanto obesidad como la diabetes tardía actuaban como mediadoras frente al cáncer de páncreas. Este estudio justificó la necesidad de desarrollar nuevas propuestas metodológicas que permitieran comprender mejor la compleja relación entre obesidad, diabetes y cáncer de páncreas.

En concreto, primero se propuso una nueva metodología, *MRinCMA*, como una extensión del análisis de mediación utilizando la *Aleatorización Mendeliana*, que permite obtener resultados insesgados dentro del contexto de mediación. Utilizando técnicas de simulación, se mostró que el sesgo de las estimaciones obtenidas con *MRinCMA* era menor que el observado cuando se utilizaron los modelos de ecuaciones estructurales, considerados como método de referencia. Se aplicó esta nueva propuesta para evaluar el efecto de la obesidad en cáncer de páncreas, considerando la diabetes tardía como mediadora y no se encontró evidencia de causalidad.

En segundo lugar, se planteó una extensión de la aleatorización *Mendeliana* que permite estratificar la población sin introducir sesgo *collider.* Considerando de nuevo métodos de simulación, se mostró que el nuevo enfoque da lugar a estimadores insesgados y a la vez facilita la identificación de la heterogeneidad entre los diferentes estratos de la población.

De manera global, se ha mostrado la importancia de aplicar correctamente métodos de inferencia causal y las diferencias con modelos clásicos de regresión. Además, se han propuesto extensiones para la *Aleatorización Mendeliana* que pueden ser usadas en otros escenarios biomédicos.

Chapter 1: Introduction

Chapter 1

1.1. Causal inference language and concepts

Causal inference is the use of statistical concepts, statistical assumptions, and diagrams to answer causality-related questions. First, it is important to distinguish the difference between association and causation. In classical association analysis, the main objective is the assessment and inference of the relationship between two or more random variables and to study the aspects of the joint distribution of those observed variables, using, for example, means, variances, regression analysis, etc (Hernán and Robins 2020; Pearl 2010a). In other words, we can say that two variables are independent when the value of one variable does not provide any information on the second variable. On the other hand, to be able to infer causality, where the objective is to study how a variable is affecting the behaviour of another variable, association analysis is not sufficient. In general terms, causality implies association (i.e., if a variable causes another, indisputably, they are associated), but association does not imply causation: two variables may be associated but do not present a causal relationship between them (Hernán and Robins 2020; Pearl 2010a). Causal language and inference can be applied when the study aims to know if a variable changes when the other variable is modified.

In many fields, the use of causal inference methodology has been increased in the last decades, due to the property of inferring causal relationships among two or more variables using experimental and observational data (Imbens and Rubin 2015).

Specifically, in epidemiological and clinical settings, researchers have been interested in studying the causal effect of a treatment (i.e., medication) or an exposure (i.e., smoking), on the development of some health outcome (i.e, headache, or cancer development, respectively). To infer causal relationships between the treatment/exposure and the outcome, classical statistical analysis is not adequate since it will only provide information regarding the association between the variables. In the associational analysis, the association between the variables would not only represent the relationship between the treatment/exposure and the outcome, but it would also represent the association constellation of all the other variables that may influence that specific relationship.

In recent years, the "evidence-pyramid" has been proposed as a potential study-design hierarchy (Murad et al. 2016; Rosner 2012), upon on the potential evidence that each study-design may offer. The pyramid of evidence classifies as the best source of evidence the systematic reviews and meta-analysis (of both experimental and observational studies), followed by randomized controlled trials (RCTs), observational cohort studies, observational case-control studies, and case-reports, Figure 1.

Systematic reviews and meta-analysis are the standard methods to assess the causal relation between health interventions and events/conditions. For instance, Ren et al. (2015) conducted a systematic review and meta-analysis to evaluate the survival benefit of a preoperative induction treatment vs. combined chemoradiotherapy, in patients with lung cancer. They included information from three randomized controlled trials. This study design, where all the relevant studies are considered regardless the impact factor of the journal in which they are published, provides very strong evidence about the effect of the treatment on the outcome.

However, systematic reviews and meta-analyses have been criticized for the poor quality of the included studies (Esterhuizen and Thabane 2016; Ganeshkumar and Gopalakrishnan 2013), and, more important, based on our concern about causation, systematic reviews and meta-analysis may include papers in which the aim is not to assess causality, but only association.

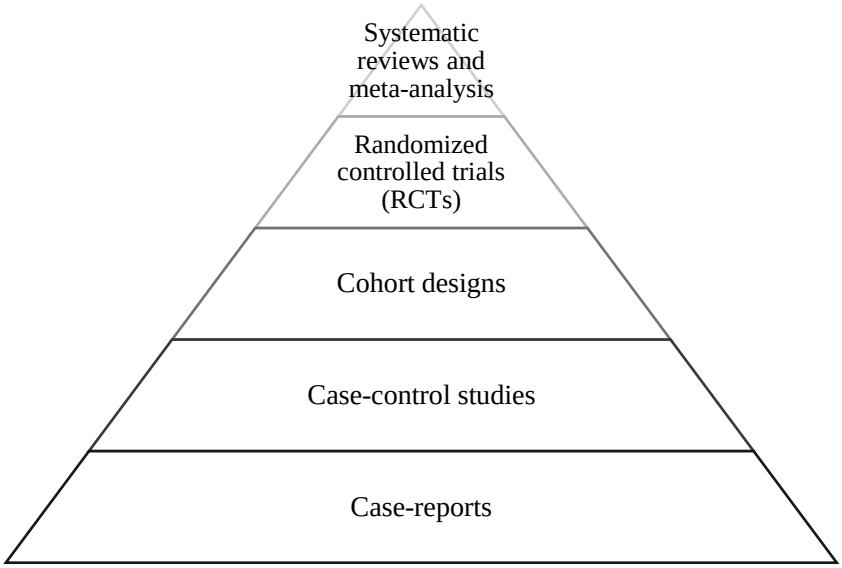


Figure 1. Pyramid-evidence of study-designs

On the other hand, when the main objective is to study the causal effect of a phenomenon on an outcome, the randomized experiments (i.e., RCT) are the gold standard. In these studies, the treatment/exposure is randomly assigned. In clinical practice, RCTs are very common when researchers are interested in the assessment of a new treatment for a specific disease. For instance, Zenke et al. (2021) evaluate the 10 year survival rate for patients with lung cancer who have received chemotherapy with thoracic radiotherapy. In this study, patients were randomly assigned to receive one of three different treatments: A, B, or C. Based on this randomization, the observed differences among groups would be only due to the allocated treatment, which means that association and causation are equivalent. Therefore, the main advantage of running a RCT is the robust evidence that it provides in terms of causal inferences, which is not easy to observe in other studies that do not include the randomization assignment. However, despite this strong advantage, the RCT presents some disadvantages that will be explained in detail in Section 1.1.4.

Fortunately, several studies have shown the similarities between RCT and observational studies (Faraoni and Schaefer 2016; Hannan 2008). Other authors such as Anglemyer, Horvath, and Bero (2014) assessed the impact of the study design on the effect estimated. They evaluated several methodological reviews which compared RCT, and observational studies and the results showed little evidence of differences on the effect estimates between both designs across all reviews, suggesting that observational studies may be a valid alternative when RCT are not suitable.

However, some important differences between RCT and observational studies should be pointed out. For example, Schad et al. (2018) carried out research similar to Zenke et al. (2021) study but under the framework of the observational studies They evaluated the effect of *Viscum album L.* jointly with chemotherapy as a treatment for lung cancer patients. In this case, patients were non-controlled, and non-randomized to any treatment, therefore the obtained results only refer to associational analysis as they were not obtained using the proper statistical methodology to infer causal relationships. Moreover, observational studies have to rely on untestable assumptions (e.g., no unmeasured confounding, or instrument validity) to obtain results as good as RCT.

Unlike the randomized allocation in RCT, in the clinical practice, the decision of giving or not a treatment to a patient is usually made by a clinician, being influenced by the clinical characteristics and health status of the individual and the physician preferences. Analogously, in the epidemiological setting, it would difficult or even unethical to randomize an individual to a specific damaging exposure (e.g., smoking).

For that reason, there is an increasing demand for causal inference methods for observational studies, since such methods allow researchers to make causal statements using observational data when RCTs are not applicable. For example, Kato et al. (2020) applied the Propensity Score technique in observational studies to assess the effect of chemotherapy in lung cancer patients, offering a valid solution to answer causal questions as the treatment was not randomized. As we will discuss in this thesis, the Propensity Score technique may provide causal conclusions if it is applied correctly.

With these examples, we illustrated the wide range of scenarios that can be used to answer a unique clinical or epidemiological question, causality, and how choosing a specific study-design may affect final results. Moreover, we also showed the similarity between randomized experiments (i.e., RCT) and observational designs and the emerging use of causal inference methodology in observational studies to answer causal questions. Nevertheless, important assumptions about the data and the causal structure must be made before starting with the causal inference analysis (see Section 1.1.2.).

1.1.1. Identification of causal effects

The first step in the causal inference analysis is the identification of the causal effects. In this section, the individual and average causal effects will be defined using the Potential Outcome notation, as described by Rubin (1974).

Let us start by considering a binary variable X (e.g. treatment with aspirin), which presents two different values: $x=1$, and $x=0$ according to the assignment or not to the treatment, the outcome Y (e.g. $y=1$, and $y=0$ according to the presence/absence of a headache) and a set of covariates, Z, for $i=1,\ldots,n$ individuals. To know whether X has a causal effect on Y, first, the individual i should be treated with level $x=1$; second, the same individual i, should be treated with the level $x=0$. If there is a difference in both outcomes after receiving (or not) the treatment, that would indicate that X has a causal effect on Y.

Definition: Potential Outcome

Rubin defined the potential outcome, $Y(x)$, as the value that the outcome Y would have taken if X had been set to the treatment level x. More specifically, the potential outcome $Y(1)$ is the value that Y would take if the treatment level was set to $x=1$. The potential outcome $Y(0)$ is the value that Y would take if the treatment level was set to $x=0$. For a specific individual, i:

$$Y_i(x)=\begin{cases} Y_i(1) & \text{observed potential outcome if } x=1 \text{ for the ith-individual} \\ Y_i(0) & \text{observed potential outcome if } x=0 \text{ for the ith-individual} \end{cases}$$

The potential outcome refers to the possibility of observing either one of these two outcomes depending on the treatment value that the individual i truly received. To study whether X has a causal effect on Y in the individual i, we should compare both potential outcomes, under the different levels of X.

Definition: ICE

The individual causal effect (ICE) for an *i-th* individual, would be the comparison of both potential outcomes in the same individual:

$$\text{ICE}_i = Y_i(1) - Y_i(0) \quad i=1,2,\ldots,n$$

The identification of this effect is not possible because an individual i, in practice, will experience only one of the two potential outcomes, $Y_i(1)$ or $Y_i(0)$.

Assuming that the identification of the individual causal effects is not feasible because only one potential outcome is observed, there is the need to identify an average

causal effect (ACE), where the potential outcome observed for one individual would be considered as the potential outcome for another individual.

Definition: ACE

An average causal effect, (ACE) is the comparison of two potential outcomes, and it is obtained by comparing both distributions of $Y(x)$, as it is shown in Table 1.

$$\text{ACE} = Y(1) - Y(0)$$

Individual	Covariates *Z*	Treatment *X*=1	Control *X*=0	ICE_i	ACE
		Potential Outcomes *Y*(1)	Potential Outcomes *Y*(0)		
1	Z_1	$Y_1(1)$	$Y_1(0)$	$Y_1(1) - Y_1(0)$	
.		.	.	.	
.	.	.	.	.	
.	.	.	.	.	Comparison of
i	Z_i	$Y_i(1)$	$Y_i(0)$	$Y_i(1) - Y_i(0)$	$Y_i(1)$ and $Y_i(0)$
.	.	.	.	.	in a set of *n* units
.	.	.	.	.	
.	.	.	.	.	
n	Z_n	$Y_n(1)$	$Y_n(0)$	$Y_n(1) - Y_n(0)$	

Table 1. Causal inference notation. Adapted from Rubin (2005)

If the difference between both potential outcomes is zero (ACE=0), it would mean that there is a null causal effect, indicating that, in average, the treatment does not have a causal effect on the outcome. Otherwise, it would mean that X has a causal effect on Y.

1.1.2. Assumptions

Because both potential outcomes in practice cannot be observed and the only information available is the one provided by the data, some assumptions needed under which the causal effects can be identified are the following.

i. Assumption I: Interference between subjects

Statement: A subject's potential outcome under treatment X does not depend on other subjects' treatment values (Hudgens and Halloran 2008; Tchetgen and Vanderweele 2012).

This assumption state that two or more individuals are independent to each other and what happens to one individual will not *interfere* with the others.

ii. Assumption II: Consistency

Statement: For each individual, one of the two potential outcomes, (i.e., the potential outcome that corresponds to the treatment level that the individual truly received) is the true observed outcome: $X_i = x \rightarrow Y_i = Y_i(x)$ (Pearl 2010b; VanderWeele 2009)

iii. Assumption III: Conditional exchangeability

Statement: Conditioning on a set of covariates, Z, the treatment level x is independent of the two potential outcomes: $X \perp Y(x) \mid Z, \forall x, z$

This is the formulation of the "no unmeasured confounding" assumption. The treated and untreated participants would be exchangeable if the assignment of the treatment only depends on the covariates Z. This assumption would hold if all the possible covariates that would influence the treatment assignment are measured.

Under the assumptions of conditional exchangeability and consistency, the causal estimands in terms of potential outcomes can be rewritten in terms of observational data, also known as *identification of the effects*.

If X and Y are binary, the *marginal causal risk difference*, MCRD, is defined as

$$\mathrm{MCRD} = \Pr\left(Y(1)=1\right) - \Pr\left(Y(0)=1\right)$$

which compares the probabilities of event between both treatments.

i. Based on the Law of Total Probability, the marginal probability can be written as:

$$\Pr\left(Y(x)=1\right) = \sum\nolimits_z \Pr\left(Y(x)=1 \mid Z=z\right)\Pr\left(Z=z\right) \qquad [1]$$

ii. Based on the assumption of conditional exchangeability, this can be written as:

$$\Pr\left(Y(x)=1\right) = \sum\nolimits_z \Pr\left(Y(x)=1 \mid X=x, Z=z\right)\Pr\left(Z=z\right) \qquad [2]$$

iii. And, based on the assumption of consistency, this is:

$$\Pr\left(Y\left(x\right)=1\right) = \sum_z \Pr\left(Y=1 \mid X=x, Z=z\right)\Pr\left(Z=z\right) \qquad [3]$$

The focus of interest is no longer defined in terms of the potential outcome. Now, the causal effect can be identified because it depends on real observed data:

$$\text{MCRD} = \Pr\left(Y(1)=1\right) - \Pr\left(Y(0)=1\right) = \qquad [4]$$

$$= \sum_z \Pr\left(Y=1 \mid X=1, Z=z\right)\Pr\left(Z=z\right) - \sum_z \Pr\left(Y=1 \mid X=0, Z=z\right)\Pr\left(Z=z\right)$$

The first stage to identify the effects is to rewrite the estimand as a new one that does not involve any variables that are not in the real data and any potential outcome. Once we have the estimand in terms of observational data, we can estimate the causal effect.

iv. Assumption V: Positivity

Statement: Treated and untreated subjects are observed and measured in any categories of the covariates Z:

$$\text{if } \Pr\left(Z=z\right) > 0 \Rightarrow 0 < \Pr\left(X=x \mid Z=z\right) < 1$$

1.1.3. Directed Acyclic Graphs (DAGs)

A DAG is a diagram where the nodes (vertices) represent random variables, V, with directed edges (arrows) and no directed cycles used to representing the factorization of a joint distribution. In causal inference, DAGs are used to draw the causal assumptions regarding the relationships among two or more variables.

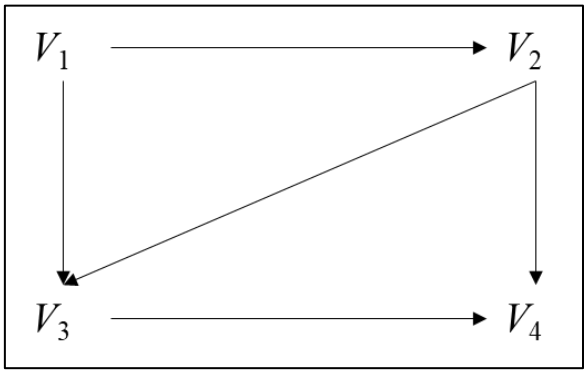


Figure 2. Example of DAG with four nodes

- A variable denoted as *parent* or *ancestor* is the node from which an arrow goes to another node. In Figure 2, V_1 is parent of V_2, and V_2 is a parent of V_3 and V_4. Analogously, V_2 is denoted as a *child* or *descendant* of V_1, and V_3 and V_4 are children of V_2, because they are nodes that are a consequence of another node. V_3 and V_4 are *descendants* of V_1 and V_2. In the same way, V_1 is an *ancestor* of V_4 and V_2.

- *Common effect*, or *colliders,* are a set of variables that are children of two other variables. In Figure 2, V_3 and V_4 are both *common effects*, or *colliders*, on the path: $V_1 \rightarrow V_3 \leftarrow V_2$ and $V_3 \rightarrow V_4 \leftarrow V_2$, respectively.
- *Common cause* is the set of variables that are parents of two other variables. V_2 is a common cause of V_3 and V_4.
- A *path* is any arrow-based route between two variables on the graph. Paths can be open or blocked according to a set of graphical rules known as D-separation rules.

Conditional on its parents (direct causes), a variable V_j is independent of any variable for which it is not a cause. Using DAGs, the lack of an arrow between two nodes indicates that these variables are conditionally independent given the other variables that have arrows into the variable. For example:

$$V_4 \perp V_1 \,|\, V_2, V_3$$

represents that variables V_1 and V_4 are conditionally independent given V_2 and V_3.

The D-separation rules are criteria used to decide whether two variables are *directional* separated, or also named, *d*-separated. These definitions were described by (Pearl 1995):

- Rule nº1: If there are no variables being conditioned on, a path is blocked if and only if two arrowheads on the path, collide at some variable (i.e., to a collider variable). In Figure 3A, the path from V_1 to V_3 is open because there is no collider variable between V_1 and V_3. However, in Figure 3B, the path between V_1 and V_2 is blocked through the collider V_3.
- Rule nº2: Any path that contains a non-collider that has been conditioned on (represented using a squared or box around the variable) is blocked. In Figure 3C, the path is blocked after conditioning on V_2 (there is no possible pathway from V_1 to V_3 because it is blocked by V_2).
- Rule nº3: A collider that has been conditioned on, does not block the path, Figure 3D.
- Rule nº4: A collider that has a descendant that has been conditioned on, does not block the path, Figure 3E.

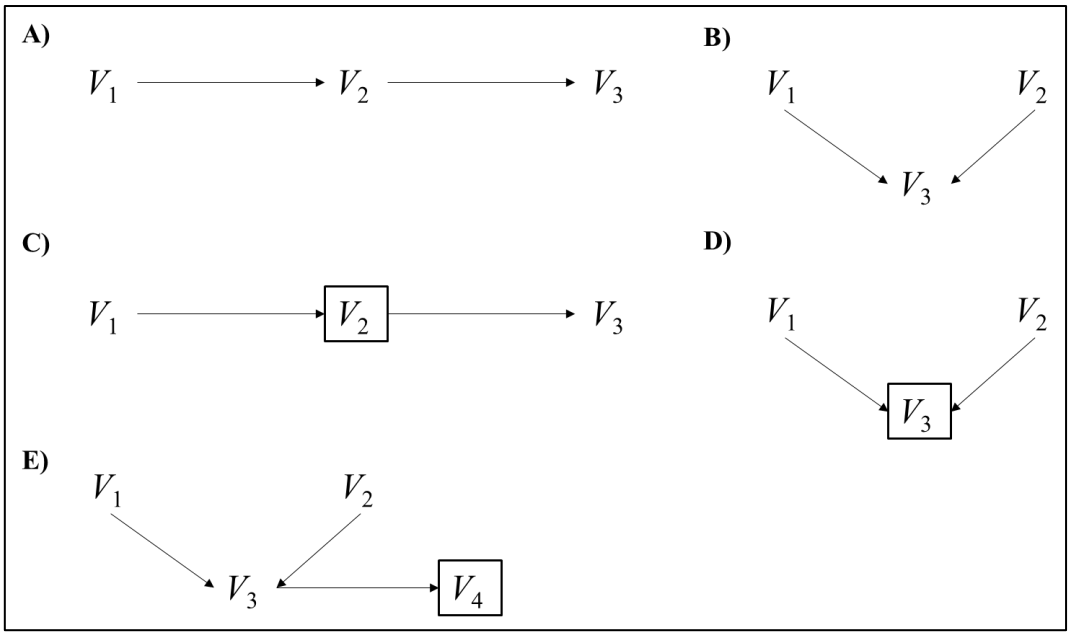


Figure 3. Examples of DAG. A), D) and E) represent open paths, B) and C) represent blocked pathways

To summarize, the three key concepts for using DAGs are:

- ✓ A path is blocked if, and only if, there is a non-collider that has been conditioned on (Figure 3C) or if there is a collider variable that has not been conditioned on (Figure 3B).
- ✓ Two variables are D-separated if all the paths between them are blocked.
- ✓ Two variables are marginally independent if they are D-separated without conditioning.

To use DAGs to infer causality and considering them as causal DAGs, some assumptions are needed:

- ✓ The lack of an arrow between two nodes would be interpreted as the absence of a causal effect between both variables, meaning that a variable does not cause a second variable.
- ✓ All common causes must be represented in the diagram (in practice, all the variables must be measured).
- ✓ Any variable is a cause of its child.

1.1.4. Types of studies: randomized controlled trials and observational studies

In clinical and public health settings, RCTs are the gold standard for the evaluation of the intervention effect, either being a treatment or a preventive action (Del Rey Calero et al. 1996).

These types of studies are characterized for being:

i) Experiments, where the investigators decide which is the intervention and outcome of interest.
ii) Prospective, where the information is collected at the study conception or while the study is ongoing.
iii) Randomized, where the assignment to the intervention is randomized. Based on this randomization, the distribution among individuals that received and did not receive the intervention would be balanced: the characteristics of both groups would be similar and the only difference between both groups would be the allocated intervention, considering that the other characteristics are homogeneous among both groups.

With this design, the potential outcomes can be considered equivalent to the observed outcomes, and additionally, the association between the intervention and the outcome would be interpreted as causation.

However, RCTs also present several drawbacks that make them inappropriate in some specific contexts. In general, to conduct them is expensive and time consuming. Moreover, ethical concerns regarding the randomization have an important role in the use of RCT, and some specific clinical questions cannot be answered using this design. For example, it would not be ethical to randomize individuals to the smoking group or to living in neighbourhoods with more air pollution. Furthermore, the selection criteria defined in RCTs, in general do not consider specific subgroups of the population (children, pregnant women, elderly patients, etc.), leading to a non-representative sample of the general population.

Observational studies are another type of study designs where the exposure or treatment assignment is not at random (Hernán and Robins 2020; Del Rey Calero et al. 1996). These types of studies are very common, and they present several advantages compared to the RCT. With this design, both costs and follow up times can be reduced and since it allows to include broader patient profiles, findings obtained can be extrapolated to a larger population. Besides in several settings, random allocation of the treatment/exposure may not be possible. For example, in clinical studies it is the clinician who decides which treatment is more appropriate to each individual of the study, based on specific clinical or demographic information. In the epidemiological setting, the exposure of interest cannot be randomized, and it will be observed by the researchers.

However, considering that baseline factors can influence the exposure or treatment assignment, many biases can arise from observational studies. The main reason is

the potentially unbalance distribution of these characteristics between groups, therefore any observed difference in the outcome, may be due to the exposure/treatment received, but also due to other covariates. Consequently, when the aim of a study is to analyse the causal effect of a treatment or exposure on an outcome, observational studies can present some limitations.

1.1.5. The biases in observational studies

A systematic bias regards to the association between two variables that does not arise from the direct association among them. These biases are consequences of certain causal structures, which can be represented using causal DAGs and can be classified according to their causal structure. An arrow from the exposure X to the outcome Y would represent a direct causal effect. This is the effect of interest to be estimated. In this thesis, we focused on two sources of systematic bias, confounding and collider bias, described in more detail in Sections 1.1.5.1 and 1.1.5.2.

In a randomized experiment, where there are no common causes thanks to randomization, the arrow from X to Y, as shown in Figure 4A, would represent causation. On the other hand, when there is no arrow from X to Y, Figure 4B, would mean that there is no causal effect of X on Y, but based on the common causes C, an association would be detected between X and Y, as described in the Section 1.1.3. ("*when a non-collider is not conditioned, it will open the path between X and Y through C*"). In general, an association between *X* and *Y* is formed by two components: i) the true causal effect of X on Y- Figure 4A, and ii) the common causes- Figure 4B, also known as *confounding* variables, which generates confounding bias, Figure 4C.

1.1.5.1. Common causes: Confounding bias

Confounding is the bias that results from the presence of common causes of X and Y, or also defined as the presence of an *open backdoor path* between X and Y. It is possible to block this open backdoor path conditioning to the *confounder* (Greenland 2003; Hernán and Robins 2020).

A backdoor path is the path which connects X and Y without using any of the arrows that leave from X to Y. For instance, in Figure 4C, the path from X to Y through C is a backdoor path. The *backdoor path criterion* states that a causal effect of X on Y can be identified if there is enough data to block all potential backdoor paths between X

and Y. Conditioning on the common causes, C, would eliminate the confounding bias, blocking the backdoor path that goes from X to Y through C. As the d-separation statement claimed, conditioning in a non-collider variable blocks the path, as shown in Figure 4D.

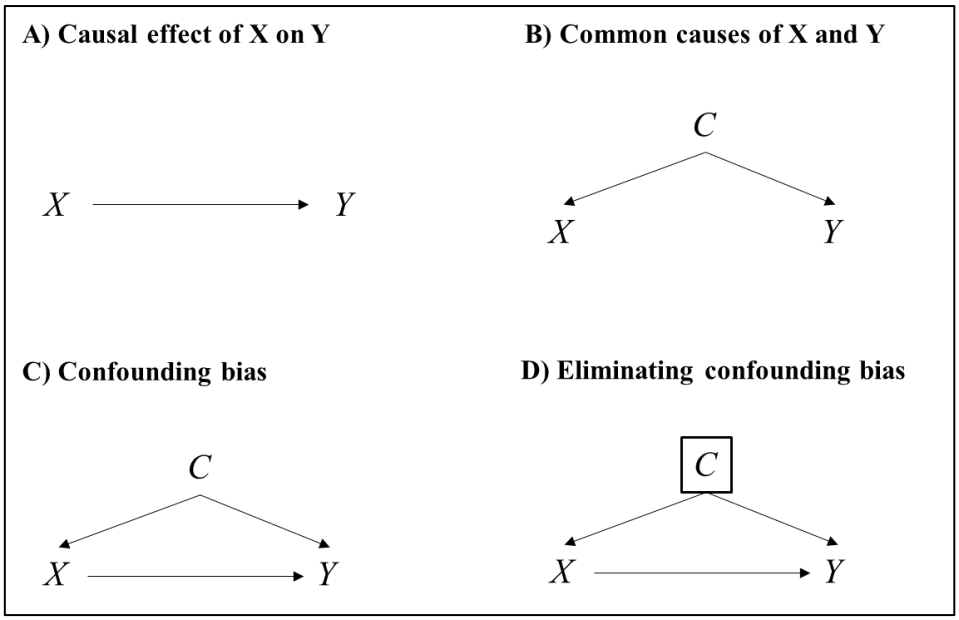


Figure 4. Definitions of confounding bias through causal DAGs

To eliminate or reduce confounding bias, it is crucial that all the variables considered as confounders, which are needed to block the backdoor path, are known, and measured. To find the proper confounder variables, it is important to consider the information available about the causal structure and the causal DAG. Two main definitions of what confounder is, can be used in practice:

- "Conventional": a variable C is a confounder if it is associated with X, C is associated with Y regardless of X, and C is not in the causal pathway from X to Y.
- "Structural": is the existence of a confounder variable which is part of an open backdoor path from X to Y, based on the causal structure and knowledge from the data. This allows identifying the confounders that are needed to be adjusted to block the open backdoor path.

Both "structural" and "conventional" definitions are valid to identify confounders, and both rely on the adjustment of the confounders to eliminate confounding bias.

1.1.5.2. Conditioning on common effects: Collider stratification bias or selection bias

Collider bias may occur when conditioning on a common effect of the treatment/exposure and the outcome, Figure 5 (Greenland 2003; Hernán et al. 2004; Hernán and Robins 2020):

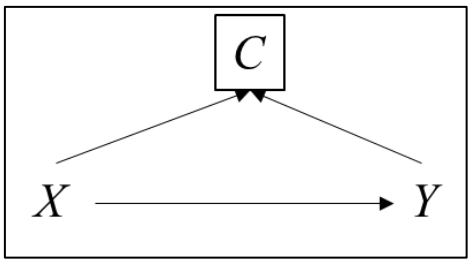


Figure 5. Example of collider bias adjusting for common effects, using causal DAGs

Conditioning on a common effect (i.e., collider variable) of X and Y opens the backdoor path that goes from X to C and Y, generating bias. Selecting a specific stratum of a collider variable creates an association between X and Y even when there is no association among them. Depending on the variable that C represents, collider bias can be known as collider stratification bias or selection bias. Collider stratification bias may occur, for instance, when the aim is to study the causal effect of X on Y but restricted to specific values of a third variable, which is a collider. On the other hand, selection bias is the bias that is due to the process by which individuals are recruited.

1.2. Statistical methods for causal inference

In general, causal inference methods are used to control and adjust for confounding bias, since it is one of the most common biases in observational studies, as it is mentioned in Section 1.1. The best way to control for confounding would be using a randomized experiment, where the treatment/exposure is randomly assigned. The possible common causes (i.e., confounders) of the treatment/exposure and outcome will not affect the treatment assignment because of the randomization. Due to the difficulties in applying RCTs in many clinical and epidemiological scenarios, observational studies are used as valid designs to answer causal questions. Nevertheless, using observational studies without the proper methodology would lead to biased and non-reliable results, because of the possible biases that could arise due to confounding. For that reason, during the last decades, there has been observed

an increased use of the causal inference methodology in observational studies, which provide valid and unbiased results in terms of causal inference interrogations.

The developed methodology has been divided into two main types, according to whether the confounding variables are collected or not in the study and if they can be used in the analysis (i.e., if the conditional exchangeability assumption holds). If all the potential confounders are collected in the study, they can be used to block the backdoor path between the treatment/exposure and the outcome, avoiding potential confounding bias. However, if there is a possibility of having potential unknown/unmeasured confounders that are not collected in the study (i.e., unmeasured confounders), this criterion is no longer valid. According to the type of confounder available in the data, the methods that rely on observable and measured confounders are the Propensity Score (*PS*), and the Causal Mediation Analysis (CMA). On the other hand, if there is a possibility of having unmeasured confounders, Instrumental Variable approaches, such as Mendelian Randomization (MR), are used.

Figure 6 represents the evolution in the number of publications for the following terminologies for the last ten years: CI = Causal Inference, CMA = Causal Mediation Analysis, MR = Mendelian Randomization, and *PS* = Propensity Score. The search was done using the *PubMed* platform (November 2021) in which each terminology has been searched independently. The Propensity Score was the most used approach, followed by the Mendelian Randomization. As it will be described in the following sections, the Propensity Score can be used with clinical and epidemiological variables, while the Mendelian Randomization requires information about genetic variants, which are not always available and easy to obtain.

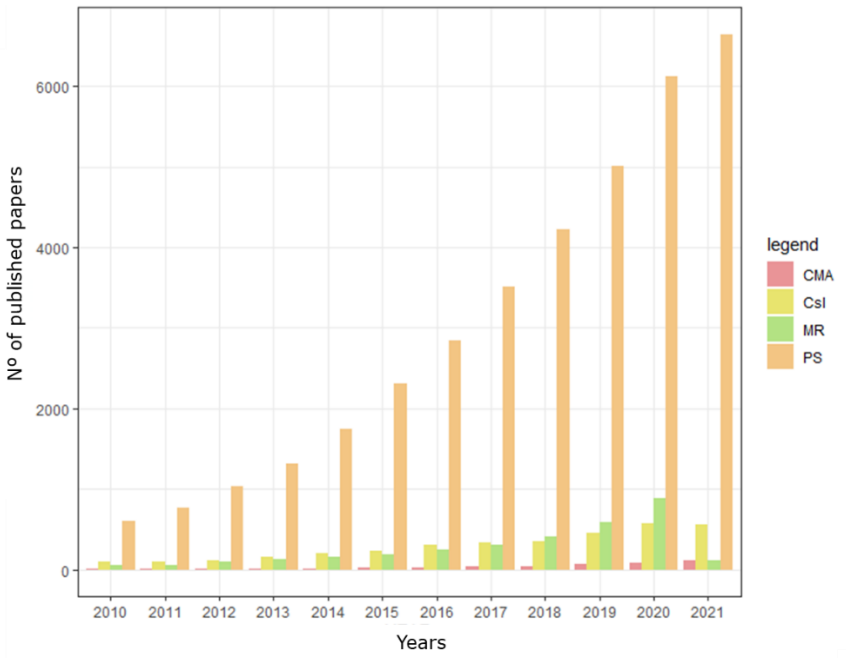


Figure 6. Number of publications of the different causal inference methods, where CsI: Causal Inference, CMA: Causal Mediation Analysis, MR: Mendelian Randomization, PS: Propensity Score. Information extracted from PubMed.

These results suggested that the Propensity Score, defined for the first time in 1983 by Rosenbaum and Rubin (1983), has been widely used in the last decade, mainly due to accessibility of the variables needed (it would require only clinical and epidemiological information). On the other hand, Mendelian Randomization was described for the first time in the early 2000's, by Davey Smith and Ebrahim, (Smith and Ebrahim 2003, 2005) because of the emerging use of genetic information in the epidemiological practice, considering that the first read of the human genome was performed in 2003. Moreover, only a few years ago, (Imai et al. 2010, 2011; Pearl 2012b) the extension of the mediation analysis in the causal inference framework was described. On the other hand, Mendelian Randomization (MR) and Causal Mediation Analysis (CMA) are novel methodologies which development and extension are still far away from Propensity Score.

1.2.1. Methods that rely on the backdoor criterion: measured confounders

To eliminate or mitigate confounding, assuming that all the confounders are measured, we could adjust for them to block the open backdoor path, as it is described in Section 1.1. One methodology widely used in the clinical settings is the Propensity Score.

1.2.1.1. Propensity Score (*PS*):

Rosenbaum and Rubin (1983) defined the Propensity Score as the probability of being treated, based on the confounders. For instance, considering a binary treatment X, that is modelled using logistic regression, and a vector C of confounders C_1 to C_n, the Propensity Score is defined as:

$$PS = \Pr\left(X_i = 1 \mid C_{i1}, \ldots, C_{in}\right) = \frac{e^{\left(\alpha_0 + \alpha_1 C_1 + \ldots + \alpha_n C_n\right)}}{1 + e^{\left(\alpha_0 + \alpha_1 C_1 + \ldots + \alpha_n C_n\right)}} \quad [5]$$

This Propensity Score models the probability for each individual of being treated (X=1) based on the $C_1, \ldots, C_n$ confounders. One of the key properties of the Propensity Score is being a *balancing* score, where the distribution of the confounders' vector, conditioning on the PS, is similar between treated and untreated individuals. A pair of individuals with similar PSs, would be also similar in their covariate values. To correctly use the Propensity Score, two main assumptions must be hold:

Propensity Score assumptions

i. Treatment assignment is independent of the potential outcomes, conditional on the observed confounder variables.

$$Y\left(x\right) \perp X \mid C, x = 0,1$$

This can be also written in terms of the PS:

$$Y\left(x\right) \perp X, C, x = 0,1 \rightarrow Y\left(x\right) \perp X \mid PS, x = 0,1$$

ii. Any subject has a non-zero probability of receiving the treatment:

$$0 < \Pr\left(X = 1 \mid C, Z\right) < 1$$

Box 1. Propensity Score assumptions

Once the *PS* is obtained, it can be used in several forms (Austin 2011):

- *Matching* subjects, is the most common way to use the PS. It creates a new subset of data generating pairs of subjects that present similar PS values. The objective of matching is creating a balanced dataset, where the confounding variables are balanced between treated and untreated subjects, eliminating (or reducing) the possible differences that could exist between the two groups. The matching *PS* is often compared to a RCT, in which the matched individuals are

similar in terms of their baseline characteristics and any observed difference would be due to the treatment assignment.

- *Inverse Probability Treatment Weighting* (IPTW) method, generates a pseudo-population without confounding bias. In this method, weights are assigned to individuals of the study based on the inverse of their conditional probability of receiving the treatment level that they actually received, which is estimated by the *PS*. These weights are defined as:

$$W_i = \frac{X_i}{PS_i} + \frac{1-X_i}{1-PS_i} \qquad i = 1,...,n \qquad [6]$$

with X_i taking value x=1 if the ith individual received the treatment or x=0, other case, and PS_i, is the Propensity Score for the ith individual. Then weights can be rewritten as:

$$W_i = \begin{cases} \frac{1}{PS_i} & \text{if } X_i = 1 \\ \frac{1}{1-PS_i} & \text{if } X_i = 0 \end{cases} \qquad i = 1,...,n \qquad [7]$$

The causal estimate is obtained using a weighted least squares (WLS) regression model considering the individuals' W_i as their respective weights. In this context, where a pseudo-population is used, the association estimate obtained would be the causal estimate.

- *Stratifying* the analysis, this method consists of dividing the whole sample in different subsamples based on the *PS* score values. Rosenbaum and Rubin (1984), demonstrated that 5 subsamples are the optimal scenario and that with this number of subsamples, the bias can be reduced at least in a 90%.
- *Including the PS as a covariate,* this method fits a multivariate regression model where treatment and PS are considered as explicative variables.

Independently of the approach used, the average causal effect (ACE) can be estimated as it is explained in Section 1.1.1.

1.2.1.2. Causal Mediation Analysis (CMA):

Mediation analysis is a specific technique used to study the effect of a treatment/exposure, X, on an outcome, Y, when there is a third variable, called the

mediator, M, which is present in the path between X and Y. Based on these variables, three effects can be defined:

i) Direct effect, is the effect of X on Y and is not influenced by the mediator M.

ii) Indirect effect, defined as the effect of X on Y that goes through M.

iii) Total effect, is the sum of both direct and indirect effects, defined as the effect of X on Y considering all potential pathways, Figure 7.

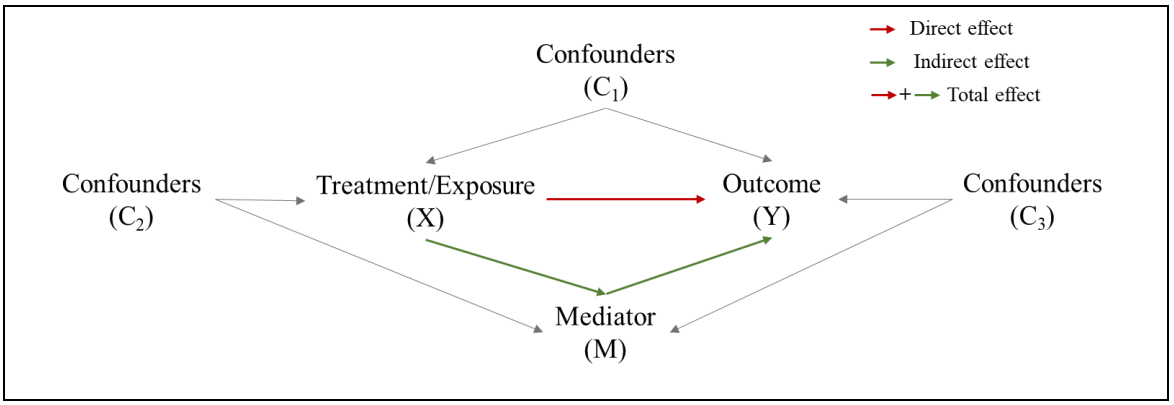


Figure 7. Mediation Analysis diagram. Direct effect is represented with the red arrow from *X* to *Y*, the indirect effect as the two green arrows, from *X* to *M* and from *M* to *Y*. The total effect is the sum of direct and indirect effect

There are two traditional approaches to run mediation analysis, defined by Baron and Kenny (1986), known as the difference and the product methods.

- The difference method, consists of fitting two regression models in which the outcome Y is the dependent variable. Suppose that Y, M and C, are continuous variables and X is a binary variable referring to the exposure variable.

The first regression model is defined as:

$$Y = \alpha_0 + \alpha_1 X + \alpha_2 C + \varepsilon_Y \quad [8]$$

where α_1 represents the total effect of the risk factor X on the outcome Y.

The second regression model is similar to [8], but now the mediator variable is included as covariate in the model:

$$Y = \beta_0 + \beta_1 X + \beta_2 M + \beta_3 C + \varepsilon_Y \quad [9]$$

where β_1 is the direct effect.

With these two equations, the indirect effect will be estimated through the difference $\hat{\alpha}_1 - \hat{\beta}_1$.

- The product method is also based on two equations. The first model is equal to [9] and the second regression model is:

$$M = \theta_0 + \theta_1 X + \theta_2 C + \varepsilon_M \quad [10]$$

The direct effect is β_1 as in the difference method, while in this case the indirect effect will be estimated using the product of $\hat{\beta}_2$ and $\hat{\theta}_1$, and the total effect is the sum of direct and indirect effects.

Definition: NDE

Robins and Greenland (1992) and Pearl (2001) formalized, in causal terms, the notion of mediation. The natural direct effect (NDE) is defined as:

$$\text{NDE} = E\left[Y(1, M(0)) - Y(0, M(0))\right] \quad [11]$$

where $M(x)$ are the potential values of M under different assignments of the exposure and $Y(x,M(x))$ are the potential values of the outcome under different assignments of exposure and mediator. The NDE is interpreted as the change in the outcome Y if the exposure X were set at level $x=1$, versus $x=0$, but for each individual when M is maintained at the level it would have taken in the absence of exposure (VanderWeele 2016).

Definition: NIE

The natural indirect effect (NIE) is:

$$\text{NIE} = E\left[Y(1, M(1)) - Y(1, M(0))\right] \quad [12]$$

The NIE is interpreted as the average change observed in the outcome Y if the exposure were controlled to $x=1$, but the mediator changed from the level it would have taken if $x=0$ versus the level it would have taken if $x=1$.

Definition: TE

The total effect (TE), as in Baron and Kenny terminology, is the sum of NDE and NIE, defined as the overall change of the outcome Y for a change in the exposure of x=0 to x=1.

In order to identify the estimands of interest (i.e., NDE, NIE, and TE), four assumptions regarding no unmeasured confounding of X, M and Y are necessary, using notation presented in Figure 7:

Causal Mediation Analysis assumptions

i) All the confounder variables between exposure and outcome (C_1) are measured

ii) All the confounder variables between exposure and mediator (C_2) are measured

iii) All the confounder variables between mediator and outcome (C_3) are measured

iv) None of the mediator-outcome confounders are affected by the treatment/exposure (VanderWeele 2016; VanderWeele and Vansteelandt 2009)

These assumptions are similar to the "unmeasured confounders" assumption described in Box 1 for the Propensity Score method, which states that all the possible confounders in the study have to be known and measured in order to block all the potential backdoor pathways between variables of interest. These assumptions are very important in CMA, and possible violations of them could lead to invalid results (VanderWeele 2016; VanderWeele and Vansteelandt 2009).

In epidemiological and clinical settings, it is quite common that the variables of interest are binary. For instance, X may be measuring whether the patient has received or not a treatment (i.e., yes, no), and the same would occur for M (diabetes yes/no) and Y *(*event yes/no*)*, respectively. For that reason, in CMA, authors proposed special definitions of the causal effects of interest, which differ from the continuous case. In terms of regression modelling, the outcome Y is regressed using a logit transformation:

$$\log\left(\frac{P_Y}{1-P_Y}\right) = \beta_0 + \beta_1 X + \beta_2 M + \beta_3 C + \varepsilon_Y \quad [13]$$

where $P_Y = \Pr(Y = 1)$ represents the probability of event. The mediator M is regressed as in Eq. [10] but using a logit transformation:

$$\log\left(\frac{P_M}{1-P_M}\right) = \theta_0 + \theta_1 X + \theta_2 M + \theta_3 C + \varepsilon_M \quad [14]$$

with $P_M = \Pr(M = 1)$.

When the outcome is rare (Valeri and VanderWeele 2013; VanderWeele 2016) and the i)-iv) assumptions (Box 2) hold, the effects can be defined as:

$$OR_{NDE} \cong \exp\left(\hat{\beta}_1\right) \quad [15]$$

$$OR_{NIE} \cong \exp\left(\hat{\theta}_1 \cdot \hat{\beta}_2\right) \quad [16]$$

The total effect (TE) is the product of both. Standard errors can be obtained using the delta method or by bootstrapping techniques.

1.2.2. Methods that do not rely on the backdoor criterion: unmeasured confounders

1.2.2.1. Instrumental Variables

The instrumental variable (IV) approach was defined in the econometrics fields and it is used in the structural equation models context. The utilization of instrumental variables (IVs) to answer causal questions was described by Angrist, Imbens, and Rubin (1996). This statistical approach allows to estimate properly the causal effect of a treatment/exposure on an outcome when there are unmeasured confounders (i.e., the backdoor path cannot be blocked). It is noteworthy that Angrist and Imbens were recently awarded (October 2021) with the Economy Nobel Prize[1] due to their contribution to the IV approach.

[1] https://www.nber.org/news/joshua-angrist-david-card-and-guido-imbens-awarded-nobel-prize

Suppose that X is the treatment or exposure, Y is the outcome of interest, and C a vector of (measured and/or unmeasured) confounders, Figure 8. A variable is defined as an Instrumental Variable (IV) if it satisfies the following conditions:

Instrumental Variables assumptions

i. The IV is associated with treatment X, (relevance condition): $IV \not\perp X$

ii. The IV does not affect the outcome Y except through X, (exclusion restriction): $IV \perp Y \mid X, C$

iii. The IV is not associated with any confounders, (marginal exchangeability): $IV \perp C$

Box 3. Instrumental Variables assumptions

Considering these assumptions, only the first one can be verified, supposing that the association between IV and X can be validated, but it will be difficult to know whether there are other pathways where there would be an association between IV and Y, or C, respectively.

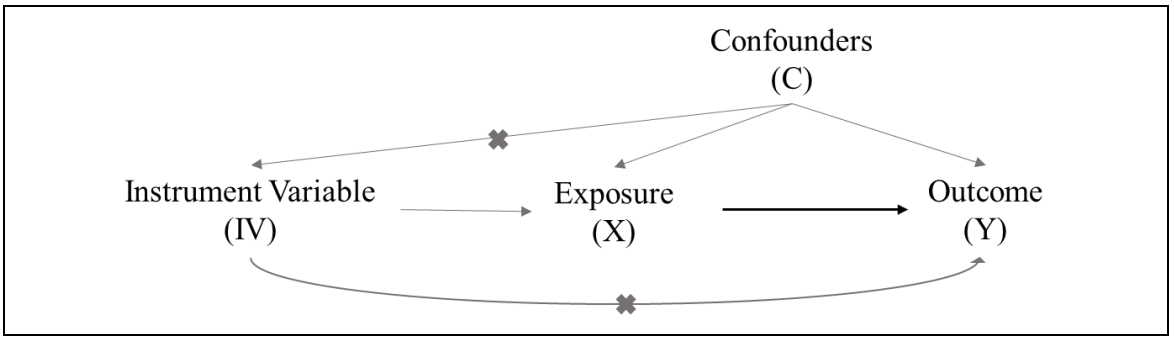


Figure 8. Instrumental Variable (*IV*) diagram, where the three *IV* assumptions are represented with the arrows from *IV* to *X*, *C* to *IV*, and *IV* to *Y*, respectively.

The IV and Y are independent if and only if, there is no causal effect of X on Y (i.e., no arrow from X to Y):

$$IV \perp Y \Leftrightarrow X \perp Y \mid C$$

Therefore, testing the association of IV on Y implies testing the causation of X on Y conditioning on C. However, even if the IV satisfies the assumptions described in Box 3, these are not sufficient to allow the identification of a causal effect of X on Y.

In the case of binary *IV* and assuming that the causal effect of X on Y, β, is the same for all individuals:

$$Y_i(1) - Y_i(0) = \beta, \ \forall i \qquad [17]$$

Therefore, ACE is β and it can be estimates as:

$$\hat{\beta} = \frac{\hat{E}\left[Y \mid IV = 1\right] - \hat{E}\left[Y \mid IV = 0\right]}{\hat{E}\left[X \mid IV = 1\right] - \hat{E}\left[X \mid IV = 0\right]} \qquad [18]$$

This ratio is known as the Wald estimate and represents the causal effect of X on Y, where the numerator in [18] denotes the effect of IV on Y, and the denominator is the effect of IV on X (S Burgess et al. 2017; Imbens and Angrist 1994).

For binary treatment and outcome, X and Y, the same notation can be used, in terms of probabilities:

$$\hat{\beta} = \Pr\left[Y(1)\right] - \Pr\left[Y(0)\right] = \frac{\Pr\left[Y = 1 \mid IV = 1\right] - \Pr[Y = 1 \mid IV = 0]}{\Pr\left[X = 1 \mid IV = 1\right] - \Pr[X = 1 \mid IV = 0]} \qquad [19]$$

Note that with this formulation, there is no need to adjust for the confounders C, because of the conditions on IV, (Box 3).

This estimand (i.e., the causal effect) can also be derived fitting two simple linear regression models, the so-called two-stage-least-squares (2SLS):

- The first model is:

$$X = \alpha_0 + \alpha_1 IV + \varepsilon_X \qquad [20]$$

- The second model is:

$$Y = \beta_0 + \beta_1 \hat{X} + \varepsilon_Y \qquad [21]$$

  where $\hat{X}$ is the prediction vector obtained in [20].

Model [21] yields the estimated parameter, $\hat{\beta}_1$, which is the estimator of the causal effect of X on Y, we emphasize that unlike in [8] and [9], in this case it is not necessary to adjust for C. This method can also be considered if either X or Y or both are binary, replacing linear models [20] or/and [21] by logistic regression models. However, non-linear equations may not guarantee that residuals from the second model [21] are uncorrelated with the IV, which could lead to misinterpretation of the results (Burgess and Thompson 2015a).

In addition to the assumptions described in Box 3, and for the point estimation of the causal effect, one of these two parametric assumptions is needed:

- The homogeneity assumption, which assumes that the causal effect of X on Y is constant for all individuals in the population (Swanson and Hernán 2013).
- The monotonicity assumption (Sheehan and Didelez 2020; Small et al. 2017; Swanson et al. 2015; Swanson and Hernan 2018), which assumes a monotonic relationship between the IV and X, (any change in X from changing IV should be in the same direction for all individuals).

Specifically, if we denote by $X(IV)$ the potential outcome of X under the IV assignment, this monotonicity condition can be expressed as:

$$X_i(0) \leq X_i(1) \ \forall i \text{, or } X_i(0) \geq X_i(1) \ \forall i$$

which implies that the effect of IV on X is in the same direction for all individuals in the study.

Under the monotonicity assumption, an IV estimate represents a local average treatment effect (also known as complier average causal effect). The compliers are the subgroups of the population whose value of X is influenced by the IV. Under the homogeneity assumption the IV estimate can be interpreted as an average causal effect, since assumes a constant causal effect in the population.

1.2.2.2. Mendelian Randomization

The Mendelian Randomization (MR) is the use of genetic variants as instrumental variables to study the causal effect of an exposure or risk factor on an outcome and is widely used in epidemiology (Burgess, Timpson, et al. 2015; Lawlor et al. 2008). The instrumental variable is built from one or more genetic variants, also called Single Nucleotide Polymorphisms ($SNPs$). These $SNPs$ are genetic variations in the four nucleotides in the DNA (i.e., cytosine, adenine, guanine, and thymine) which can occur in any subject, and can differ between individuals, Figure 9.

The genetic instrument in any Mendelian Randomization study can be define as i) a single genetic variant, ii) more than one $SNPs$ simultaneously, or iii) using an allele score.

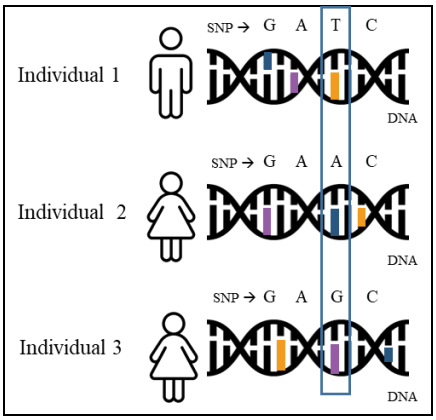


Figure 9. Example of three individuals with different *SNP* values. *SNP*: Single Nucleotide Polymorphism, G: Guanine, A: Adenine, T: Thymine, C: Cytosine

The allele score is a weighted score that is obtained as the sum of two or more $SNPs$ weighted by their respective effects in the exposure and it is define as (Burgess and Thompson 2013):

$$G_i = \sum_{j=1}^{m} \beta_j SNP_{ij}, \quad i = 1,...,n \qquad [22]$$

where G_i is the allele score for the ith individual, m is the total number of $SNPs$ used; SNP_{ij} is the value of the jth SNP in the ith individual, which can take values 0, 1, 2; β_j is the effect of the SNP_j on the exposure and is derived using a linear/logistic regression model where the exposure is regressed on the SNP.

The causal estimate, $\hat{\theta}$, can be obtained using the Wald ratio estimate:

$$\hat{\theta} = \frac{\hat{\beta}_{Y|G}}{\hat{\beta}_{X|G}} \qquad [23]$$

where $\hat{\beta}_{Y|G}$ and $\hat{\beta}_{X|G}$ are obtained from models [24] and [25]:

$$Y = \beta_Y + \beta_{Y|G} G + \varepsilon_Y \qquad [24]$$

$$X = \beta_X + \beta_{X|G} G + \varepsilon_X \qquad [25]$$

G is either a single SNP or the allele score defined in [22]. Afterwards, the 95% confidence interval is constructed considering the standard error as:

$$se(\theta) = \frac{S_{Y/G}}{\beta_{X/G}} \qquad [26]$$

Analogously, this causal estimate can also be obtained using the Two-Stage-Least-Square (2SLS) or applying the Wald ratio estimate as described in Section 1.2.2.1.

As for the general IV approach, three assumptions must be considered before running a MR study:

Mendelian Randomization assumptions:

i) The genetic instrument, G, is associated with the exposure, X:

$$G \not\perp X$$

ii) The genetic instrument, G, does not affect the outcome, Y, directly, only indirectly through the exposure, X:

$$G \perp Y \mid X$$

iii) The genetic instrument, G, is not associated with any measured or unmeasured confounders, C:

$$G \perp C$$

In practice, the second and third assumptions can be violated in presence of horizontal pleiotropy, linkage disequilibrium (LD), or population stratification. Some limitations can occur when running a MR analysis, which are the following (Stephen Burgess et al. 2017; Zheng et al. 2017):

- Weak instrument bias, which could arise when the selected genetic variants are not strongly associated with the risk factor of interest.
- The horizontal pleiotropy (i.e., genetic confounding), that occurs when a genetic variant is associated with the outcome Y in different pathways than the risk factor of interest.
- Another form of pleiotropy is the vertical pleiotropy, also known as *mediation*, which occurs when a genetic variant is associated with multiple traits that are in the same pathway.
- The linkage disequilibrium (LD) is a type of confounding that occurs when there is a correlation between two or more genetic variants, especially in those genetic variants that are closer in the same chromosome.
- The population stratification bias occurs when subgroups within the total sample have different genetic ancestries.

According to the available information of the participants, Mendelian Randomization can be divided in:

- **Individual level data:** This approach is used when all the information from the individuals is collected and available.
- **Summary data:** This method uses summarized genetic associations that are already published, (i.e., the genome-wide association study UKBiobank[2], GWAS catalog[3], TCGA[4]).

Based on the sources used, Mendelian Randomization can be defined as:

- **One sample MR:** As its name suggests, a unique dataset is used for all participants. For this case, either the 2SLS or the Wald estimate can be considered to estimate the causal effect.
- **Two sample MR:** Two datasets are used for the analysis, one for the risk factor and one for the outcome.

In this way, there are four different cases of MR:

1. **One sample, individual level data**: Similar to classic IV analysis, where both Wald estimate and TSLS can be used.
2. **Two sample, individual level data:** More uncommon in practice, due to the difficulty of having two complete sources.
3. **Two sample, summary data:** This is the standard source in MR analysis, and where different MR approaches can be used.
4. **One sample, summary data:** Similar to the two-sample summary data case, where all the MR approaches can be applied.

Another common MR approach, a part from the Wald ratio and 2SLS, is the **Inverse variance weighted (IVW)**, described by Bowden et al. (2015), which uses a meta-analysis to combine the Wald estimates obtained from different $SNPs$. The IVW causal effect, is estimated as:

$$\hat{\beta}_{IVW} = \frac{\sum_j \hat{\beta}^2_{X|G_j} \hat{\theta}_j \sigma^{-2}_{Y_j}}{\sum_j \hat{\beta}^2_{X|G_j} \sigma^{-2}_{Y_j}} \qquad [27]$$

[2] https://www.ukbiobank.ac.uk/

[3] https://www.ebi.ac.uk/gwas/home

[4] https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga

where $\hat{\beta}_{X|G_j}$ are the G_j-X association estimates, $\hat{\theta}_j$ are the causal estimates on the *j*th *SNP* and σ_{Y_j} is the standard error of $\hat{\theta}_j$, for $j=1,\ldots,m$.

The same estimand can also be obtained from a weighted linear regression:

$$\hat{\beta}_{Y|G_j} = \beta_{IVW}\hat{\beta}_{X|G_j} + \varepsilon_{Y_j} \quad \text{where } \varepsilon_{Y_j} \sim N\left(0, \sigma_{Y_j}\right);\ j = 1,\ldots,m \qquad [28]$$

The IVW is in general used in the two sample summary data settings, since it does not require individual level data, as the 2SLS.

Robust Mendelian Randomization methods:

Several robust methods were proposed in order to avoid potential biases due to IV violations:

- **MR- Egger regression:** This approach considers the violation of the *IV3* assumption, Box 3, which may occur when there is directional pleiotropy. Contrary to the IVW method, the MR-Egger regression does not constrain the intercept to zero. The MR-Egger model is defined as:

$$\hat{\beta}_{Y|G_j} = \beta_{0E} + \beta_{1E}\hat{\beta}_{X|G_j} + \varepsilon_{Y_{E_j}} \quad \text{where } \varepsilon_{Y_{E_j}} \sim N\left(0, \sigma_{Y_j}\right);\ j = 1,\ldots,m \qquad [29]$$

  If the intercept term equals zero, then $\hat{\beta}_{1E}$ would be the same as $\hat{\beta}_{IVW}$. However, if the intercept term differs from zero, there is evidence of directional pleiotropy and $\hat{\beta}_{1E}$ will be a consistent estimate of the causal effect, under de InSIDE assumption (i.e. correlation between $\hat{\beta}_{X|G_j}$ and $\hat{\beta}_{Y|G_j}$ is zero) (Bowden et al. 2015), when both the sample size and the number of SNPs increase (Burgess and Thompson 2017). In this case the intercept would be the average pleiotropic effect of a genetic variant 0.

- **Weighted median:** It is used in the specific scenario where there are invalid instruments (i.e., genetic variants with pleiotropy effects). If 50% or more of the genetic variants are considered valid instrumental variables, then the IV estimates for these variants would be consistent estimates of the causal effect. The median of all the IV estimates based on the individual genetic variants will be a consistent estimate. More efficiently, it is better to use a

weighted median, in which the more precise genetic variants receive more weight than others, (Bowden et al. 2016).

- **Multivariable Mendelian randomization (MVMR):** It is used when the genetic variants are associated with two or more correlated exposures, and the aim is to study how two or more exposures affect the outcome. Furthermore, this technique is used also when ii) and iii) assumptions in Box 3 are violated. The MVMR would use genetic variants that are associated with the multiple exposures but do not affect the outcome. With this method, possible causal effects per each exposure can be estimated (Burgess and Thompson 2015b).

1.3. Data sources and applied examples

1.3.1. Intensive Care Unit (ICU) database

Intensive Care Units (ICU) are specialized units where critical patients receive constant care and they are monitored by a highly specialized team. The first hours in ICU, jointly with the treatment decision made by the physician, are crucial to increase their survival probabilities. Moreover, ICU patients are heterogeneous: they are admitted due to different health diseases and conditions, and they may present different baseline characteristics, which will influence the physician's decisions. Altogether, ICU patients pose a challenge in their analysis, especially for studying causal treatment effects.

In the second chapter of this PhD thesis, a database from several Intensive Care Units (ICU) in Spanish hospitals was used. This dataset included information between 1998 and 2016, where the material was collected every six years (i.e., 1998, 2004, 2010, 2016). During the stay in the ICU, the patients were followed up and several variables on the clinical conditions were collected. The initial number of patients was 3,857, with a total number of clinical variables of 590. We restricted the database to the years 2010 and 2016, which were the years where specific variables such as mechanic ventilation were collected. Restricting to patients with no missing values for some variables of interest, the final dataset used had 2,335 individuals and it includes, around 30 variables of interest.

The complexity of using this database relies on the strong confounding bias that could be present between many variables, mainly because the nature of the study was

observational. Moreover, considering that the information was related to ICU patients, the clinical decision based on the patients' health status was crucial and the final analysis could be biased. Furthermore, in this type of studies, it would not be ethical to randomize any treatment considering the health condition of many patients. For that reason, causal inference approaches such as the Propensity Score took into consideration all the possible confounding variables and could be easily applied in this type of context where the randomization was not a valid option, but it still was interesting studying the causality among specific variables.

1.3.2. Pancreatic cancer and the PanGenEU study

Pancreatic cancer is the seventh most common cause of cancer related mortality worldwide, with an overall 5-year survival rate of 9% (Allemani et al. 2018). China and the United States of America (USA) are the countries with the highest pancreatic cancer incidence worldwide and it is expected that it will be the second most common cause of death by cancer in the USA by 2030 and Europe (Rahib et al. 2014). Although pancreatic cancer does not present the highest incidence among cancer types, it is shown that it presents the highest mortality rate (Figure 10).

The most common type of pancreatic cancer (>90%) is the infiltrating ductal adenocarcinoma (PC). This in an aggressive disease, with a serious resistance to chemotherapy and radiotherapy treatments. Moreover, most of the patients with PC have been diagnosed in an advance stage of the disease with cancer metastasis, being one of the reasons that there is not yet a reliable screening for the early detection of this disease.

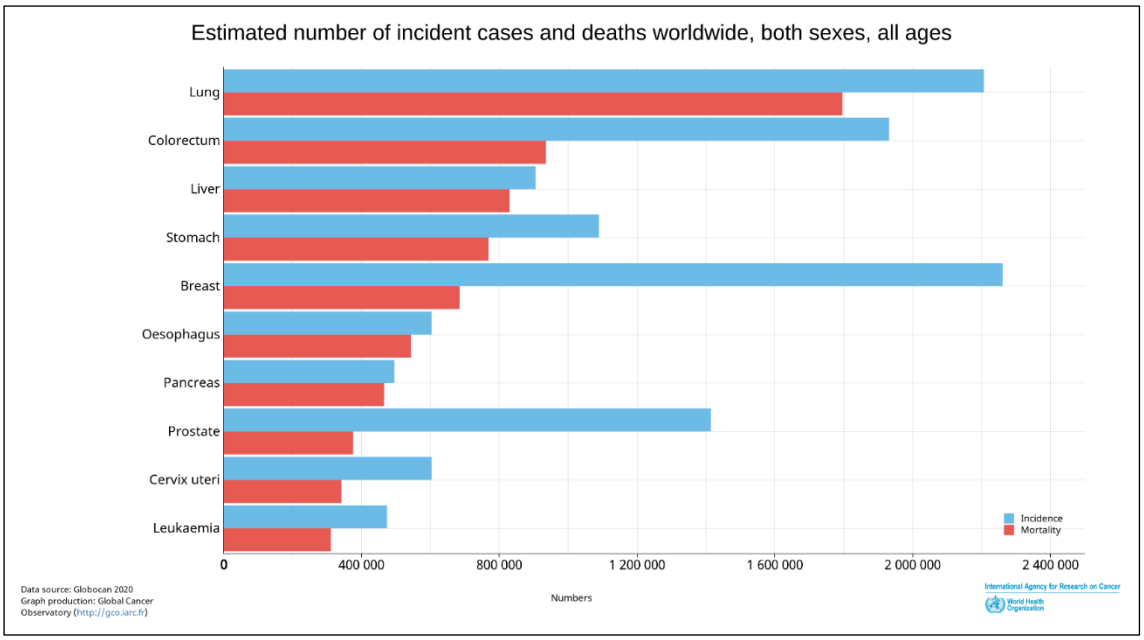


Figure 10. Pancreatic cancer incidence and mortality across the 10 most deadly cancer types[5].

The aetiology of pancreatic cancer is complex, and it is still not well characterized. It is known that some risk factors are related to lifestyle or the environment, whereas specific heritable factors may contribute to pancreatic cancer development. Based on the World cancer report, as non-heritable factors, smoking, heavy alcohol consumption, race, age, obesity, type-2 diabetes mellitus, chronic pancreatitis, or non-O blood group are defined as important risk factors (Huang et al. 2021; Maisonneuve and Lowenfels 2015).

As for type 2 diabetes mellitus (T2DM) and obesity, while both are considered risk factors, they present a complex interaction because both conditions are related, and they often coexist, Figure 11. Moreover, T2DM can be split into two subtypes according to the time of diagnosis: long-standing diabetes mellitus (LSDM) if it is diagnosed two years before PC diagnosis, while a new-onset diabetes mellitus (NODM) is defined when the diabetes diagnosis occurs less than two years before pancreatic cancer diagnosis. The NODM is also known as the type 3c diabetes, and it can be considered as an early sign of pancreatic cancer.

[5] Figure obtained from GLOBOCAN https://gco.iarc.fr/

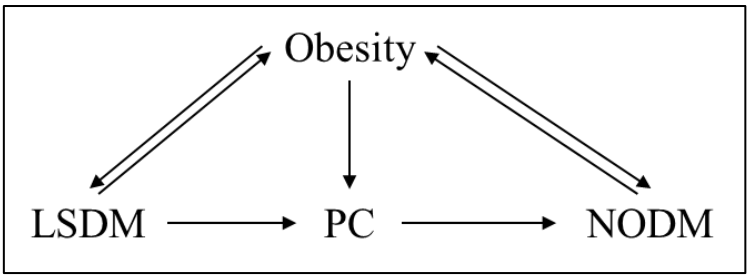


Figure 11. Interplay between pancreatic cancer, Type 2 Diabetes Mellitus subtypes, and obesity. LSDM: Long standing diabetes mellitus, PC: Pancreatic ductal adenocarcinoma, NODM: New onset diabetes mellitus.

Pancreatic cancer is a deadly disease, with poor and unknown characterized risk factors. Therefore, it is crucial to study the epidemiological and genetic causes of this disease, to be able to identify risk factors that could be used for early prevention and detection. Until the day, the most common approach to study which risk factors are associated with pancreatic cancer risk was based on associational approaches, while there are not many studies using causal inference methodology.

In this PhD thesis, we focussed on two main important risk factors, the T2DM subtypes (LSDM and NODM) and obesity, due to their coexistence. Only few recent Mendelian Randomization studies analysed the causal effect of T2DM and obesity on pancreatic cancer disease (Carreras-Torres et al. 2017; Lu et al. 2020; Yuan et al. 2020). However, none of them differentiated between LSDM or NODM subtypes. We hypothesize that new insights regarding the complex relationship between pancreatic cancer, T2DM subtypes and obesity can arise using Mendelian Randomization and Causal Mediation Analysis.

To carry out this analysis, we used the resource of the PanGenEU (the European Study into Digestive Illnesses and Genetics), which is a hospital-based case-control study of pancreatic cancer conducted in six European countries (Spain, Germany, Ireland, United Kingdom, Italy and Sweden) and 28 centres. It is designed to evaluate environmental and genetic factors associated with pancreatic cancer, from 2007 to 2014, where 2,018 cases and 1,540 controls were recruited, and 1,162 cases and 752 controls were considered to extract genetic information. A standardized epidemiological questionnaire including self-reported socio-demographic and anthropometric data, family history of cancer and in particular also PC, medical history (i.e., chronic pancreatitis, diabetes, and others) that involves regular use of specific medication, and lifestyle behaviours (i.e., smoking and alcohol habits) was administered by trained personnel in a face-to-face interview, generating more than 300 variables of information.

1.3.3. Bladder cancer epidemiology and the UK Biobank

According to the World Cancer Report, bladder cancer (BLCA) is the 9th most common cancer worldwide and the 13th most common cause of death by cancer worldwide. The incidence and mortality rates are higher in North America, Europe, North Africa, Middle East, and Australia. The 72% of the new cases occurred in areas with high levels of human development. The mean age of BLCA diagnosis is around 67 years, being rare in persons aged 40 years or younger. Moreover, BLCA is 4 times more frequent in men than women, probably due to the smoking habits, and occupational exposures (Malats and Real, 2015).

Smoking is one of the most important causes of bladder cancer, yielding a 2-6 fold increased risk in smokers compared with the non-smokers. More specific, there are differences according to the type of tobacco: black tobacco smokers present higher risk of being diagnosed with BLCA than blond tobacco smokers. Considering smoking as a risk factor for bladder cancer, it is important to highlight the difference that smoking individuals present on their weights in comparison to the non-smokers. It is known that smoking and bodyweight are associated (Munafò et al. 2009; Piirtola et al. 2018), since smokers trend to present lower bodyweight levels than former or non-smokers.

In order to better disentangle the causal effect of smoking on bladder cancer disease, we attempted to differentiate the causal effect based on different bodyweight levels, to identify specific subgroups that present higher risk of bladder cancer. These groups would benefit from early screenings or clinical interventions. Using the proper Mendelian Randomization extension, we were able to identify those specific subgroups that present higher risk of bladder cancer.

For this PhD project, we used the resources from the UK Biobank[6] (Sudlow et al. 2015), which is a large and detailed prospective study with more than 500,000 individuals, with age between 40-69 years, recruited in 2006 and 2010 in 22 different centres of the United Kingdom (UK). This study collected the information regarding phenotypic and genotypic characteristics about the participants, and longitudinal follow-up for a wide range of health-related outcomes. For this specific study, a total of 367,643 individuals and 160 variables were considered.

[6] https://www.ukbiobank.ac.uk/

1.3.4. Simulation studies

An important source used in this PhD project are the simulations databases considered for the Chapters 5 and 6, respectively. Simulations studies are the most valid approach to assess and validate any new methodological proposal (Morris et al. 2019). An important advantage of using simulation studies is the possible assessment under different values and parameters considered.

In Chapter 5, to evaluate potential mediation effects using two genetic instruments, we simulated several datasets under different correlation values, effect sizes, variable types, and sample sizes, based on the PanGenEU information.

In Chapter 6, to assess the impact of collider bias, we generated datasets considering different variable types and parameters but keeping a fixed sample size of n=10,000.

1.4. Hypotheses, objectives, and thesis organization

1.4.1. Hypotheses

Due to the increased use of the causal inference methodology in the recent years, especially in the medical and epidemiological field, in this thesis it is hypothesized that:

1. The Propensity Score is a proper approach for evaluating treatment effects in clinical scenarios, while the Mendelian Randomization, and Causal Mediation Analysis are valid methodologies to answer epidemiological questions. These approaches lead to unbiased estimates even in the presence of measured or unmeasured confounding; however, in several scenarios, they may present some limitations which requires further methodological developments.
2. By extending the causal inference approaches, the disentanglement of causal mechanisms can be improved while offering a novel methodological solution. Here, I hypothesize that extended Causal Mediation analysis jointly with Mendelian Randomization yields a valid alternative to estimate the causal effect of interest, without any source of bias.
3. By extending the Mendelian Randomization, I propose a new approach that allows the stratification of the population without introducing collider bias. This new extension may offer new insight regarding high-risk population subgroups that would benefit from early interventions.

1.4.2. Objectives

The main objective of this thesis is to develop two new methodologies for the correct estimation of the causal effect in different clinic-epidemiological scenarios and simulation settings. More specific, to develop and extend the Causal Mediation Analysis using the genetic information provided by the Mendelian Randomization, in a new approach called *MRinCMA* and, to propose an extension of the Mendelian Randomization to correct for potential collider bias that can arise when the analyses are stratified by a collider variable.

Specific objectives are:

1. To apply the Propensity Score in different clinical scenarios to study the effect of neuromuscular blockers on delirium, and to study the causal effect of the interruption of sedation on the health status in ICU patients.
2. To apply the Mendelian Randomization and Causal Mediation Analysis in an epidemiological real study, to analyse the causal mechanism between T2DM-subtypes, BMI/obesity, and pancreatic cancer.
3. To further explore the causal effect of BMI on pancreatic cancer and the potential mediator role of T2DM-subtypes.
4. To study the effect of smoking on bladder cancer across different bodyweight levels.

1.4.3. Thesis organization

First, in Chapter 1, I have introduced a comprehensive set of concepts regarding the causal inference theory, the assumptions that are behind the field, and some key concepts that will be used along the thesis, including confounding bias, study designs, and a brief description of the used methodology.

Second, in Chapter 2, I present the Propensity Score methodology using a clinical scenario to study the causal effect of neuromuscular blockers on delirium, and the effect of the interruption of sedation in the health status of the patients that are in the ICU. This project entitled "Analysis of causality from observational studies and its application in clinical research in Intensive Care Medicine" has been published in the journal *Medicina Intensiva* in 2018.

In Chapter 3, I present a comparison between Propensity Score studies and Randomized Controlled Trials to assess the VTE treatment effect on patients' mortality, with a systematic review and meta-analysis. This project entitled "Comparison of All-cause Mortality Following VTE Treatment Between Propensity

Score-Adjusted Observational Studies and Matched Randomized Controlled Trials: Meta-Epidemiological Study" has been published in the journal *Chest* in 2019.

In Chapter 4, the Mendelian Randomization and Causal Mediation analysis techniques were applied in a real case scenario to analyse the causal effect of T2DM-subtypes and BMI/obesity on pancreatic cancer risk. The results of this project were published in *GUT* in 2020. The manuscript is titled "Deciphering the complex interplay between pancreatic cancer, diabetes mellitus subtypes, and obesity/BMI through causal inference and mediation analyses".

In Chapter 5, I propose a new extension of Causal Mediation Analysis using the genetic information provided by Mendelian Randomization to further study the causal mechanisms between BMI and pancreatic cancer, considering the mediator role of T2DM. The resulting work is now under revision, in a paper entitled "MRinCMA: a new proposal to address mediation analysis interrogations by using genetic variants as instrumental variables".

In Chapter 6, I propose a new approach to stratify the population avoiding collider bias in Mendelian Randomization. This paper, entitled "Stratification on a collider in Mendelian randomization without collider bias" is now under the second revision at the European Journal of Epidemiology. This paper is a joint effort with Dr. Stephen Burgess, senior researcher at the MRC Biostatistics Unit of the University of Cambridge. This is the result of a virtual internship I held from February to September 2021 to obtain the international mention of the PhD. The preprint is available in medRxiv.

https://www.medrxiv.org/content/10.1101/2021.08.17.21262178v1

In Chapter 7, I discuss the obtained results in the previous chapters, as well some limitations and possible future perspectives.

Chapter 2: Propensity Score applied in an Intensive Care Unit database

Chapter 2

In this chapter I present the study comparing the Propensity Score approaches to the classic regression models, applied in different ICU scenarios.

2.1. Background

Confounding bias is common in Intensive Care Unit databases since the treatment assignment is made based on the clinical status of the patients and any randomization is not feasible, leading to potential biased results. The Propensity Score method is commonly used in this type of scenarios, and it allows to analyse the causal effect of the treatment on the outcome reducing potential confounding bias.

2.2. Objectives

The main objective is to illustrate the different Propensity Score approaches that can be used to adjust for confounding bias in clinical settings and to compare them to classic regression models in different ICU scenarios.

2.3. Methods

The Propensity Score is the probability of receiving a treatment based on the observed baseline covariates of the patients, as it is explained in Chapter 1.

Considering this score, treated and untreated individuals can be compared, and causal conclusions can be made without any confounding bias. In this work, we applied four Propensity Score approaches, matching, Inverse Probability Treatment Weighting, Stratification and Covariate-adjustment using a Spanish database of ICU patients. The aim is to estimate the causal effect of neuromuscular blockers on delirium events and the interruption of sedation on the mortality of the patients. We compared the Propensity Score causal estimates to those obtained from a univariate and multivariate logistic regression models adjusted for the baseline covariates and confounding variables.

2.4. Results

Similar results in term of point estimates and 95%CIs across the different Propensity Score approaches and univariate and multivariate regression models were observed when estimating the causal effect of the interruption of sedation on mortality, Table 2.

Method	OR	95%CI
PS:Matching	0.64	(0.49;0.83)
PS:IPTW	0.67	(0.52;0.85)
PS:Stratification	0.68	(0.53;0.86)
PS:Covariate-adjustment	0.67	(0.52;0.85)
Univariate logistic regression model	0.64	(0.52;0.80)
Multivariate logistic regression model	0.63	(0.49;0.81)

Table 2 OR and 95%CI estimates for the interruption of sedation on mortality.

On the other hand, the Propensity Score approaches differed when studying the effect of the neuromuscular blockers on delirium compared to univariate and multivariate logistic regression, Table 3.

Method	OR	95%CI
PS:Matching	1.73	(0.55;5.97)
PS:IPTW	2.08	(0.83;5.24)
PS:Stratification	1.82	(0.74;4.42)
PS:Covariate-adjustment	2.07	(0.87;5.11)
Univariate logistic regression model	2.79	(1.26;6.17)
Multivariate logistic regression model	3.18	(1.42;7.13)

Table 3. OR and 95%CI estimates for the neuromuscular blockers on delirium.

2.5. Conclusions

Different OR values were obtained depending on the clinical scenario, number of events observed and potential confounding variables. The number of covariates that can be adjusted for depended on the number of events observed and affected the final causal estimates and approaches used. Propensity Score approaches and classic logistic regression models lead to similar results when patients were similar, while the results differed when patients presented different baseline characteristics. When the number of events is low and the number of covariates to adjust for is high, Propensity Score approaches are the best method to adjust for confounding bias.

2.6. Publication

The full manuscript of our study is included in this chapter and was published in the journal *Medicina Intensiva* (Impact Factor Journal Citation Reports 2018: 1.982; Q3; 24/33 Category *Critical Care Medicine)*

Coscia Requena C, Muriel A, Peñuelas O. Analysis of causality from observational studies and its application in clinical research in Intensive Care Medicine. *Med Intensiva* (Engl Ed). 2018 Jun-Jul; 42(5):292-300. English, Spanish. doi: 10.1016/j.medin.2018.01.002.

The citations counts of this article (21/11/21) provided by Google Scholar was 3 and by Scopus was also 3.

Med Intensiva. 2018;42(5):292–300

medicina intensiva

www.elsevier.es/medintensiva

SPECIAL ARTICLE

Analysis of causality from observational studies and its application in clinical research in Intensive Care Medicine☆

C. Coscia Requena[a], A. Muriel[a,b,c], O. Peñuelas[d,e,*]

[a] Unidad de Bioestadística, Hospital Universitario Ramón y Cajal, Madrid, Spain
[b] IRYCIS, CIBER de Epidemiología (CIBERESP), Madrid, Spain
[c] Departamento de Enfermería y Fisioterapia, Universidad de Alcalá, Alcalá de Henares, Madrid, Spain
[d] Unidad de Cuidados Intensivos y Grandes Quemados, Hospital Universitario de Getafe, Getafe, Madrid, Spain
[e] CIBER de Enfermedades Respiratorias (CIBERES), Spain

Received 7 September 2017; accepted 13 January 2018

KEYWORDS
Causality;
Clinical trial;
Observational study;
Confounders;
Propensity score;
Intensive Care;
Epidemiology

Abstract Random allocation of treatment or intervention is the key feature of clinical trials and divides patients into treatment groups that are approximately balanced for baseline, and therefore comparable covariates except for the variable treatment of the study. However, in observational studies, where treatment allocation is not random, patients in the treatment and control groups often differ in covariates that are related to intervention variables. These imbalances in covariates can lead to biased estimates of the treatment effect. However, randomized clinical trials are sometimes not feasible for ethical, logistical, economic or other reasons. To resolve these situations, interest in the field of clinical research has grown in designing studies that are most similar to randomized experiments using observational (i.e. non-random) data. Observational studies using propensity score analysis methods have been increasing in the scientific papers of Intensive Care. Propensity score analyses attempt to control for confounding in non-experimental studies by adjusting for the likelihood that a given patient is exposed. However, studies with propensity indexes may be confusing, and intensivists are not familiar with this methodology and may not fully understand the importance of this technique. The objectives of this review are: to describe the fundamentals of propensity index methods; to present the techniques to adequately evaluate propensity index models; to discuss the advantages and disadvantages of these techniques.

☆ Please cite this article as: Coscia Requena C, Muriel A, Peñuelas O. Análisis de la causalidad desde los estudios observacionales y su aplicación en la investigación clínica en Cuidados Intensivos. Med Intensiva. 2018;42:292–300.

* Corresponding author.
E-mail address: oscar.penuelasro@salud.madrid.org (O. Peñuelas).

PALABRAS CLAVE
Causalidad;
Ensayo clínico;
Estudio observacional;
Confusión;
Propensión;
Cuidados Intensivos;
Epidemiología

Análisis de la causalidad desde los estudios observacionales y su aplicación en la investigación clínica en Cuidados Intensivos

Resumen Una de las características fundamentales de los ensayos clínicos es la asignación aleatoria de un tratamiento o intervención sobre los pacientes. Esta asignación divide los pacientes en dos grupos que, aunque difieran por el tratamiento recibido, presentan unas características basales homogéneas haciendo que ambos grupos sean comparables y se pueda evaluar el efecto causal del tratamiento. Por otro lado, los estudios observacionales se caracterizan por la asignación no aleatoria del tratamiento y por lo tanto que los grupos de pacientes no solo difieran por el tratamiento recibido, sino también por otras características basales, a menudo relacionadas con la variable de intervención. En numerosas ocasiones, los ensayos clínicos aleatorizados no son factibles por razones éticas, logísticas, económicas o de otro tipo. Uno de los retos de la investigación clínica en Cuidados Intensivos debería ser aprovechar los datos que provienen de la práctica clínica habitual y analizarlos como si fueran ensayos clínicos. Los estudios observacionales utilizando métodos de análisis con índices de propensión *(propensity score)* han ido en aumento en los artículos científicos de Cuidados Intensivos. Los análisis de índices de propensión intentan controlar la confusión en estudios observacionales ajustando la probabilidad de que un determinado paciente esté expuesto. Sin embargo, los estudios con índices de propensión pueden ser confusos, y los intensivistas no están familiarizados con esta metodología y pueden no comprender plenamente la importancia de esta técnica. Los objetivos de esta revisión son: describir los fundamentos de los métodos del índice de propensión; presentar las técnicas para evaluar adecuadamente los modelos de índices de propensión, y discutir las ventajas y los inconvenientes de estas técnicas.

Introduction

In clinical research carried out in Intensive Care Units (ICUs), one of the common objectives is to evaluate the causal relationship between a given treatment or intervention (exposure) and the health outcome of a patient (death, healing, discharge from the ICU). Clinical trials are the reference research design when assessing the efficacy of a treatment in relation to the event of interest, since they reduce the risk of confounding influences or selection bias. Clinical trials involve the random assignment or allotment of a series of patients with a similar disease stage in order to assess the impact of such treatment upon the outcome. In the ICU setting, this outcome is usually defined as mortality during admission to the Unit or during the first 28 days after admission, or as the need for more aggressive treatment (e.g., tracheotomy).[1–3]

Assignment to treatment is the principal characteristic of clinical trials. The type of treatment for each patient is determined on a random basis in order to ensure that both the treated and the untreated patients present homogeneous features, thereby preventing the effect of the treatment from being confounded by the characteristics of the patients.

Although clinical trials are considered to be the highest quality studies for estimating causality, they also have some limitations: the sample size is typically of limited size and difficult to achieve; external validity is low; the application of inclusion criteria narrows down the population analyzed (since elderly patients, individuals with comorbidities or pregnant women tend to be excluded); ethical issues must be taken into account; and the duration of follow-up is limited.[4]

In the case of the ICU there are additional factors that make the designing of clinical trials in this setting particularly difficult, such as nosographic shortcomings (patients admitted to the ICU present syndromes such as for example acute respiratory distress syndrome [ARDS] instead of diseases); problems in defining adequate control groups; the concomitant use of different treatments (in many cases the intervention is not a drug but a therapeutic approach); randomization prior to treatment; or the obtainment of informed consent (which poses problems due to the moment in which consent is required).[5]

A possible solution to some of the difficulties of clinical trials is the conduction of observational studies. These represent a type of research in which treatment selection is conditioned to the baseline characteristics of the patient. The supervising physician decides the type of treatment according to the patient features. This is one of the main differences between observational studies and clinical trials.[6–9]

Observational studies have a number of advantages with respect to clinical trials: the population setting is broader; the duration of follow-up is longer; and the sample size is greater. In contrast, the fact that treatment allocation is not randomized implies that the estimation of causality is biased, and that the treated patients and untreated patients therefore differ not only in the treatment received but also in their baseline characteristics. If these baseline variables

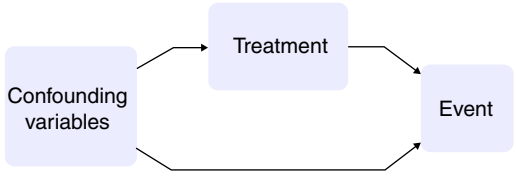


Figure 1 Representative diagram of the relationship between the confounding variables and the event in the observational studies.

were associated to the event of interest, they would act as confounders between exposure and outcome[10,11] (Fig. 1).

Statistical techniques are continuously developing, and new methods are increasingly used in clinical research to estimate causality in observational studies, considering the possible confounding factors – specifically the propensity score (PS) and marginal structural models.

Articles using PS methods have become increasingly common in the Intensive Care literature over the last 10 years (Fig. 2). Since PS methodology is still not very familiar to intensivists, the present study establishes the underlying concept and describes the best practice for applying the method. Specifically, the aims of our study are to describe: (1) the principles of PS methods; (2) the main methods for using the propensity score (matching, stratification, covariate adjustment, and inverse probability of treatment weighting [IPTW]) in an Intensive Care example; (3) statistical analysis in the presence of time dependent confounders; and (4) the strong and weak points of these techniques.

The propensity score in observational studies

According to Rosenbaum and Rubin,[12] the PS allows us to estimate the probability of assignment to a treatment, conditioned to the baseline characteristics of the patient. This is a balanced score, since treated patients and untreated patients with similar PS will have equivalent baseline characteristics. In this way we would control the confounding effect produced by the baseline characteristics.

Two conditions must be met in order to apply the PS: there must be no unmeasured confounding factors, and each subject must have a probability different from zero of receiving a treatment. If these two conditions are met, treatment allocation can be regarded as independent with respect to the results, when conditioning for the covariables.

In clinical trials the PS is known, since the probability of patient assignment to a given treatment is known (e.g., in the case of a study with two treatment arms, the probability of assignment of a patient to either arm will be 0.5). However, in observational studies this score is not known. It could be estimated through logistic regression, where the dependent variable is the treatment and the independent variables are the possibly confounding baseline covariables.

Variables to be considered in calculating the propensity score

The variables included in the PS can generate bias, modifying the variance of the estimator or error, if not correctly chosen. There is no consensus regarding the choice of such variables, since this depends on the clinical approach – defining the variables based on clinical criterion – and the statistical approach used.[13]

A given set of baseline variables can be classified into: (1) variables related to the event; (2) variables related to the exposure; and (3) variables related to both the event and to the exposure.

As we are dealing with a logistic regression model, the number of variables is determined by the total number of exposed/unexposed patients, following the rule of Peduzzi, i.e., 10 patients per event of interest (exposed/unexposed).

Simulation studies have shown that the variables related to the result (event) must be included, along with the variables showing an association to the event and to the exposure.[14] Variables only showing an association to the treatment are not to be included.

The choice of variables is a challenge in which balanced agreement between statistical findings and clinical criterion must be established.

Propensity score application methods

Once the PS has been estimated for all the patients, different application methods can be used: matching, stratification, IPTW, and adjustment of the final model, including it as a confounding covariable.

Matching

Matching is based on creating a new patient sample according to the previously estimated PS. Patients with similar PS are selected to create pairs of treated patients and untreated patients. The aim of this method is to create a new balanced sample and reduce the differences there may be between patients in the two groups. Matching allows the generation of pairs considering different aspects. In this regard, replacement could be considered, and therefore it could be decided whether an individual can be matched twice. The patient may be matched to the ''neighboring'' patient that is closer. Likewise, a caliper may be established, indicating the maximum separating distance with which two patients can be matched – the value usually being 0.25 times the standard deviation of the PS.

With this new sample, presenting a structure very similar to that of a clinical trial, we could solve the confounding problem and would be reducing the differences in the characteristics of the treated patients and controls.

In the matched sample, and based on standardized differences, we would have to check that the difference in baseline characteristics of the treated patients and the untreated patients after matching is less than 10%.[15]

The standardized differences are expressed as follows for continuous variables:

$$d = \frac{\overline{x} - \overline{x}_{\overline{T}}}{\sqrt{\frac{S_T^2 + S_{\overline{T}}^2}{2}}}$$

where $\overline{x}_T$, $\overline{x}_{\overline{T}}$ are the means of the variables in the treated and untreated patients, respectively, and S_T^2, $S_{\overline{T}}^2$ are the variances of those variables.

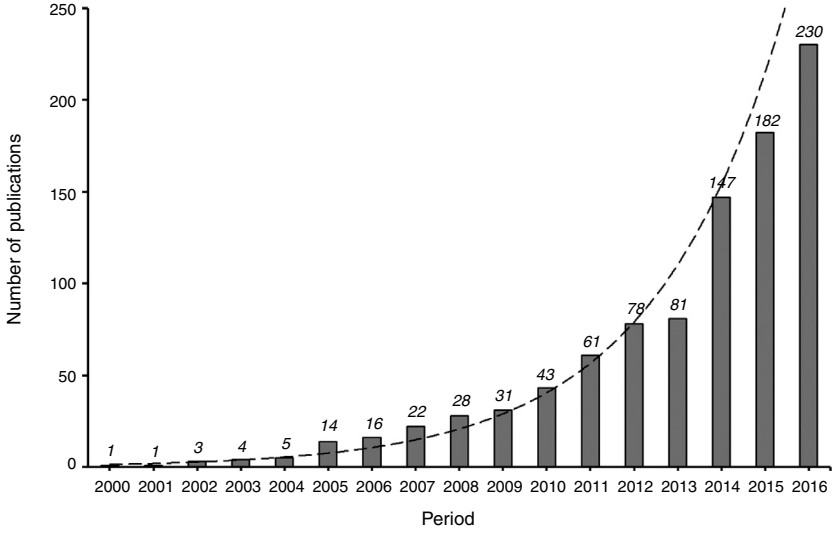


Figure 2 Flow chart of the publications in adult critical patients with use of the propensity scores.
Source: chart generated by the authors (PubMed search up until March 2017).

Likewise, we define the standardized differences for dichotomic variables:

$$d = \frac{\left(\hat{p}_T - \hat{p}_{\bar{T}}\right)}{\sqrt{\frac{\hat{p}_T\left(1-\hat{p}_T\right)+\hat{p}_{\bar{T}}\left(1-\hat{p}_{\bar{T}}\right)}{2}}}$$

where $\hat{p}_T$ is the proportion of events of a variable for the treated individuals, and $\hat{p}_{\bar{T}}$ is the proportion for untreated individuals.

The last step is to estimate the effect of the treatment upon the event of interest in the matched sample. If the response variable is continuous, we can estimate the difference between the two groups by means of the Student *t*-test or the Wilcoxon test in the absence of a normal distribution. If the response variable is dichotomic, the effect is estimated using the McNemar test.

Inverse probability treatment weighting (IPTW)
This method involves weighting estimates based on the PS, calculated as the inverse of the probability for treated and untreated patients, followed by estimation of the effect of the treatment weighted for the previously calculated weight.

Stratification
The stratification method consists of dividing the sample into several subsamples based on the PS. The number of groups is usually 5, according to quintiles or other percentiles of the estimated PS. Rosenbaum et al. indicate that bias can be lowered 90%. The effect of the treatment is estimated based on the Mantel-Haenszel (MH) odds ratio (OR).

Propensity score
The last alternative is to calculate the effect of the treatment entering the PS as covariable in the model. This method is based on the premise that the relationship between the outcome (event) and the confounding variable must be correctly specified, i.e., that it has been established correctly; the use of nonlinear models, splines or fractional polynomials could help to establish adequate models.

Comparison of the different methods

The advantage of the matching method is that once the pairs have been obtained, the new structure is similar to that of a clinical trial. In contrast, this technique requires a large sample in which there are more untreated patients than treated patients, in order to ensure that all treated patients can be matched. By matching we moreover would lose all the information corresponding to those patients that could not be matched.

The main advantage of the stratification method is that we can use the entire patient sample. Furthermore, it is a more robust technique in the event the PS has not been correctly specified in comparison with the IPTW method or covariable.

According to Deb et al.,[15] a disadvantage of the IPTW method is that the calculated weights may be unstable if there are patients with low probabilities of receiving treatment: if the weight is regarded as the inverse of the PS, those patients with a low PS will be assigned a greater weight than those with a high PS.

Considering PS as covariable of the model, we obtain a simple way to include many variables in the latter. In order to use this method, it must be checked that the relationship between PS and the outcome has been correctly specified; the assumption of a normal distribution should be confirmed if the regression model is linear. The inconvenience here is that checking the balance of the baseline variables (diagnostic balance) is more complicated than in other methods, and it has been demonstrated that more biased estimates are obtained.[15] Lastly, implementation of this method does not allow the estimation of absolute risk reduction or of the number needed to treat (NNT).

Propensity score and regression

Under certain conditions, the estimated obtained by the multivariate regression models and PS are coincident. The difference between PS and logistic regression is strongly conditioned to the number of events and to the number of covariables to adjust for.[16]

According to Harrell et al.,[17] the optimum number of independent variables that should be entered in a logistic regression model is 10 times the number of events in the sample—a condition known as the ''rule of 10''. In other words, if the number of events is 30 (deaths, for example), the optimum number of explanatory variables for estimating the probability of death would be three. Logistic regression therefore will depend on the number of events and consequently on the sample size.[18]

However, in the logistic regression of PS, the dependent variable is not the event but exposure. The number of variables for estimating exposure will depend not only on their association to the event as commented above, but also on the number of exposures there are. If for example the number of exposures is 60 (60 smokers), we should adjust PS with 6 variables.

Depending on the prevalence of the event and on the number of subjects exposed to the treatment, PS allows us to fit a model with more variables than the logistic regression model.

Time dependent confounders. Marginal structural models

The PS is a statistical technique that allows us to estimate the effect of the treatment upon the patient outcome. When the clinical confounding variables are time dependent, estimation of the score is more complex, since the values of the confounding variables vary over time.

Example. In evaluating the causal relationship between sedation (A) and the failure of noninvasive mechanical ventilation (Y), we may encounter confounding variables (L) such as for example the RASS sedation scale or certain respiratory markers such as pH or $PaCO_2$.[18,19]

Both the RASS sedation scale and the respiratory markers (pH or $PaCO_2$) are measured in different time intervals. They therefore are regarded as time dependent confounding parameters if the weighted values of the covariables can predict current treatment or if the current values can predict future results, conditioned to the treatment received in the past.

In this example, and as can be seen in Fig. 3, we seek to estimate the effect of treatment in the form of sedation (A) in both instant 0 (A_0) and in instant 1 (A_1). On the other hand, L_1 represents the confounding variables (RASS score and pCO_2) measured in instant 1, and Y represents the outcome of interest (failure of noninvasive ventilation). On studying the relationship between the treatment in instant 0 and the outcome, two causal routes appear: A_0–Y and A_0–L_1–Y. Route A_0–Y is a direct route.

However, in route A_0–L_1–Y, the relationship established between the treatment and the outcome is conditioned to the confounding variables L_1. To estimate the effect, according to the classical models, we could resort to regression

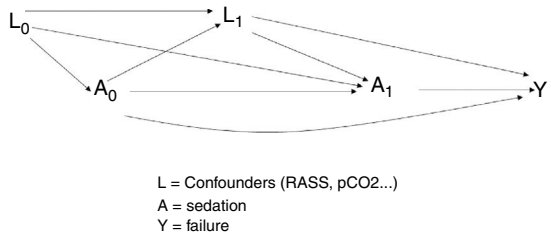


Figure 3 Representative diagram of the time dependent confounding variables in the observational studies. A_0 is the study or exposure variable; Y is the outcome of interest; L_0 represents the variables that modify the study variable as a function of time.

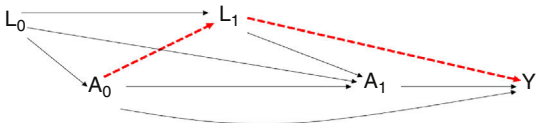


Figure 4 Representative diagram of the time dependent confounding variables in the observational studies. The possibility exists that a relationship could be established between the exposure variable and outcome simply influenced by the presence of the time dependent confounding variables (alternative route in red), and not really by the existence of a causal relationship between exposure and outcome.

without considering the confounders L_1, since they are found in the same causal chain between the effect (A_0) and the outcome (Y) (Fig. 4).

If we estimate the effect without adjusting for L_1, bias would be produced, since we would not be taking into account the causal route between A_1 and Y which is conditioned by L_1. The regression methods can pose problems of bias even in the absence of residual confounding elements. This is possibly due to the confusion created by the time dependent variables in assignment of the previous treatment.

There are a number of alternatives for estimating the effect of a treatment in the presence of time dependent confounders: marginal structural models and nested structural models.[20]

Marginal structural models[19] are considered an alternative to regression models when there is a time dependent confounding variable associated to the outcome of interest and which is also related to the treatment being evaluated. The term ''structural'' indicates a causal effect, not only a statistical association.

Instead of combined distributions, use is made of marginal distributions conditioned to the baseline variables: the term ''marginal models'' is therefore used.

Estimation of effect based on a marginal structural model

In creating the marginal structural models, the first step is to estimate PS based on logistic regression or probit models, in which the dependent variable is the treatment and the independent variables are those parameters associated with the start of treatment and/or with the event of interest.

Each patient contributes as many observations as number of time units followed (months, weeks, etc.). The variable follow-up time is to be added as covariable.

The weight of each patient and time is estimated according to the probabilities predicted by this model, based on the inverse of the probability of treatment actually received. In this context, it is assumed that there are no competing events, since each patient is considered to be followed-up on until the event of interest occurs, or the patient is lost to follow-up.

In the same way as above, a weight is estimated for each patient, for censoring.

Based on their product, we combine the weights obtained of the treatment and of the probability of censoring. With such weighting, we can simulate a sample in which treatment and censoring are independent of the measured confounders. In the event of very small probabilities, the outlier weightings are replaced by near lying percentiles[21] (1 and 99 or 2.5 and 97.5), thereby avoiding the presence of very large weighting values.

If the event of interest is a binary variable, the OR of the logistic regression between the event and treatment, weighted for the combined weight, offers a good approximation to the instantaneous relative risk of the Cox model, because the risk of events is low in all the months.

Assumptions

In order to use the above described statistical methods we need to check certain assumptions:

- The time order: exposure to treatment occurs before the event.
- Consistency: the potential result of an individual, conditioned to the observed treatment history, coincides with the true result observed.
- Positivity: the individuals of the population must show a probability of over zero of receiving the treatment in each treatment category and in each of the levels of the confounding variables. In this way, the mean causal effect of the treatment can be estimated within each subgroup of the population established by the confounding variables.
- Correct specification of the model: the model has been estimated adequately. The assumption of linearity is to be evaluated.
- Absence of residual confusion: this can be checked by means of a sensitivity analysis, since it cannot be assumed using only the data of the patients.

Application of the propensity score in patients admitted to Intensive Care Units

The fundamental advantage of the PS is that it summarizes the baseline characteristics of the investigated individuals in a single number. The score summarizes the probability of exposure with a number.

For example, if the objective of the study is to evaluate the effect of noninvasive ventilation in children admitted to the ICU in an observational study of over 30,000 patients of which almost one-third receive noninvasive ventilation as first option, this type of research question could have been addressed—with difficulties—by means of a clinical trial.[22]

This type of analysis has also been used to estimate the adjusted effect of cancer upon mortality in the ICU,[23] since this investigation can only be addressed in clinical research by means of observational studies.

We could also use these analytical techniques to estimate the effect of a complication such as weakness acquired in the ICU (WICU) upon weaning failure or patient mortality during admission to the ICU.[24,25]

We will develop this latter example step by step. The study involved is a prospective, international multicenter trial involving 4157 patients subjected to mechanical ventilation during more than 12 h. The appearance of WICU during admission was associated to an increased incidence of weaning failure and to higher mortality in the ICU. The PS was estimated with the following variables: age, SAPS II score upon admission to the ICU, main reason for starting mechanical ventilation (chronic obstructive pulmonary disease, heart failure, sepsis, acute respiratory distress syndrome and pneumonia).

In order to evaluate the impact of sedation and analgesia in patients subjected to noninvasive ventilation upon the need for endotracheal intubation, and in the presence of time dependent confounders such as the RASS (Richmond Agitation-Sedation Scale) or PCO_2, we must consider a marginal structural model. To illustrate the use of these methods comparing the different PS options, and exhibiting the results based on classical methods such as logistic regression, we will use a series of data from a multicenter study in the ICU involving mechanical ventilation.

First part

Considering as treatment the interruption of sedation and as event the death of the patient, PS has been applied to assess causality. On applying different PS modalities (matching, stratification, IPTW, and as covariable of the model) for this clinical scenario, we obtained similar estimates in terms of OR and 95% confidence intervals (95%CI). These ORs have been compared with those obtained estimating the effect without PS (Table 1).

These such comparable results are attributable to the fact that in relation to the baseline characteristics, the treated and untreated patients do not differ, and the estimated PS is therefore very similar in both sets of individuals.

Second part

If we consider a series of patients with more different baseline characteristics, the obtained estimates of the effect of treatment would not be consistent among the different methods. For example, if we only consider those patients with respiratory disease and wish to study the effect of neuromuscular blockers upon patient delirium, the different methods would yield different estimates (Table 2).

The weightings obtained from IPTW span the following values: [1.037; 12.75].

When treated and untreated patients differ, and the number of events and exposures is small, the statistical methods show significant differences.

Table 1 Relationship between mortality risk on day 28 (status day 28) and the interruption of sedation, according to different propensity analytical strategies.

	Applications of the propensity score								Without propensity score			
	Matching		IPTW		Stratification MH		Covariable		LR univariate		LR multivariate	
	OR	95%CI	OR	95%CI	OR	95%CI	OR	95%CI	OR	95%CI	OR	95%CI
Status day 28	0.64	(0.49; 0.83)	0.67	(0.52; 0.85)	0.68	(0.53; 0.86)	0.67	(0.52; 0.85)	0.64	(0.52; 0.80)	0.63	(0.49; 0.81)

CI: confidence interval; IPTW: inverse probability of treatment weighting; MH: Mantel-Haenszel; OR: odds ratio; LR: logistic regression.

Table 2 Relationship between mortality risk on day 28 (status day 28) and the interruption of sedation, according to different propensity analytical strategies.

Applications of the propensity score			Without propensity score		
	OR	95%CI		OR	95%CI
Matching	1.72	(0.55; 5.97)	Univariate	2.79	(1.26; 6.17)
IPTW	2.08	(0.83; 5.24)	Multivariate	3.18	(1.42; 7.13)
Stratification	1.81	(1.14; 7.14)			
Covariable	2.06	(0.84; 5.11)			

CI: confidence interval; IPTW: inverse probability of treatment weighting; OR: odds ratio.

In order to see how the treated and untreated patients are distributed based on the estimated PS, we examine the zone of common support, i.e., the range of common values which PS presents in both groups.

Therefore, the more similar the treated and untreated patient groups, the larger the number of patients of both groups included in the zone of common support.

If the treated and untreated patients are very similar, the zone of common support includes all the individuals, since the maximum and minimum of both PS will be similar. In contrast, if both groups are very different, the most likely result is that the number of patients entering the zone of common support will be smaller (Fig. 5 and Table 3).

In order to establish whether the order of the subjects in the matching method exerts an influence, we simulated 100 models and obtained the corresponding OR estimates. We see that when the number of variables for which PS is adjusted is high, the order of matching exerts no influence, and similar estimates are obtained. When the number of variables for which the order of matching is adjusted does exert an influence, different estimates are obtained.

Advantages and inconveniences of these techniques

Depending on the number of events and exposures involved, it has been seen that PS and logistic regression could be regarded as equivalent, yielding similar estimates when no differences are observed between treated and untreated patients, and the number of events is high.

When the number of events is low, PS allows adjustment for more variables than logistic regression, and moreover quantifies the effect of treatment.

Matching of the PS results in a structure similar to that of a clinical trial (the pairs show similar characteristics), with the disadvantage that those patients not matched are excluded from the study.

Because of the random patient allotment involved, randomized clinical trials have the advantage of being able to assess the causal relationship directly. When such studies are not viable, however, observational studies can be regarded as an alternative. In order to assess causality in observational studies, we first must solve the confounding problems they imply.

Conclusions

Clinical trials represent the best methodological design for analyzing causality in clinical research. However, certain hypotheses cannot be tested clinically due to ethical, methodological or economical limitations. In order to overcome such limitations, observational studies are able to simulate the hypothetical scenario of a clinical trial, using PS and marginal structural models to provide the desired answers. The growing interest in this new methodology makes it necessary to familiarize intensivists with it in order to facilitate its application to clinical research in the Intensive Care setting.

As an example, Delaney et al.[21] described the use of the statistical methods detailed in this article to evaluate the effect of corticosteroids upon mortality among patients with influenza A (H1N1pdm09). The ORs associated to these models ranged from 1.85 (95%CI: 1.12–3.04) in the classical multivariate logistic regression models to 1.71 (95%CI: 1.05–2.78) in the logistic regression model adjusted for PS, 1.52 (95%CI: 0.90–2.58) after matching of the PS, and 0.96 (95%CI: 0.28–3.28) in the marginal structural model adjusting for the time dependent variables. On adjusting for the time dependent variables in the marginal structural model, no association was observed between corticosteroid use and patient mortality.

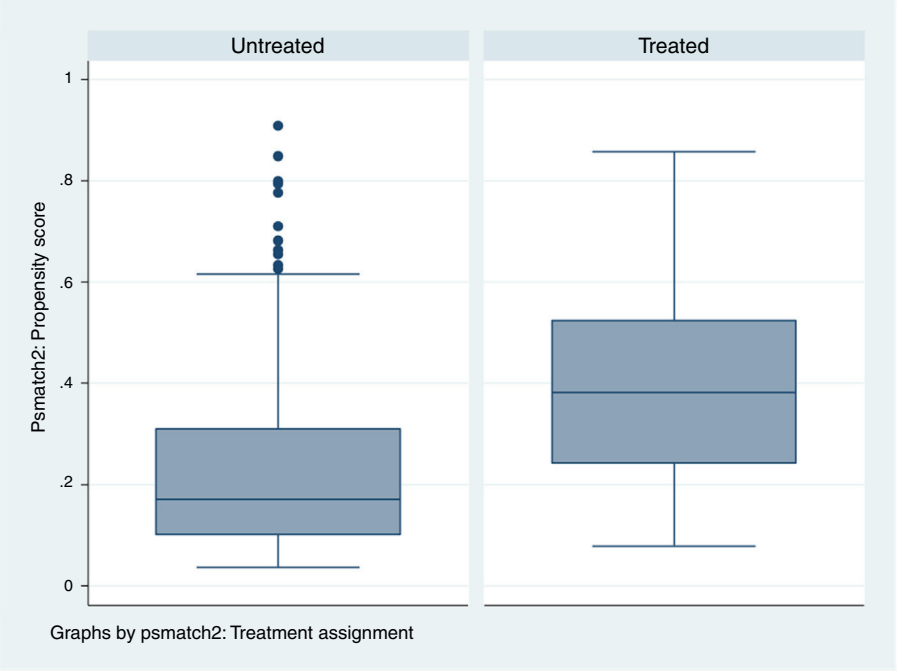


Figure 5 Box plot for comparing—in this case for observing—that the patients treated with neuromuscular blockers (NMBs) are different from the patients that have not been treated with NMBs.

Table 3 Comparison of the standardized differences of the baseline variables in treated and untreated patients, before and after matching, to confirm that the propensity score (PS) is correctly specified. If these differences exceed 10%, PS has not been adequately calculated.

	Before matching			After matching		
	NMB = yes (n = 148)	NMB = no (n = 390)	Standardized differences (%)	NMB = yes (n = 141)	NMB = no (n = 381)	Standardized differences (%)
Baseline variables						
Age (years) mean (SD)	59.53 (13.53)	65.33 (13.58)	42.79	60.53 (14.53)	58.96 (13.64)	11.17
SAPS_II (points), mean (SD)	43.74 (15.49)	46.83 (16.76)	19.15	44.99 (15.67)	43.41 (14.78)	10.37
Gender (M)	56.8	48.5	16.63	55	58.3	6.73
Invasive ventilatory support (yes)	21.6	21.3	0.59	23.3	21.7	3.98
Presence of cardiovascular failure (Yes)	68.9	40.8	58.37	67.5	68.3	1.78
Presence of renal failure (yes)	42.6	21.28	40.87	36.7	40	6.85
Presence of Hematological failure (yes)	24.3	8.97	25.75	19.2	16.7	6.52
Sepsis during mechanical ventilation	43.24	24.1	37	40.88	45.83	10.10
Duration of ventilatory support (days)	15.90 ± 13.14	8.21 ± 14.82	54.91	14.35 ± 11.60	11.97 ± 19.04	15.09

Conflict of interest

The authors declare that they have no conflicts of interest (economical, commercial or intellectual) in relation to this study.

References

1. Grasselli G, Gattinoni L, Kavanagh B, Latini R, Laupacis A, Lemaire F, et al. Feasibility, limits and problems of clinical studies in Intensive Care Unit. Minerva Anestesiol. 2007;73:595–601.
2. Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. J Educ Psychol. 1974;66:688–701.
3. Rubin DB. For objective causal inference, design trumps analysis. Ann Appl Stat. 2008;2:808–40.
4. Holmberg L, Baum M. Can results from clinical trials be generalized? Nat Med. 1995;1:734–6.
5. Villar J, Pérez-Méndez L, Aguirre-Jaime A, Kackmarek RM. Why are physicians so skeptical about positive randomized controlled clinical trials in critical care medicine? Intensive Care Med. 2005;31:196–204.
6. Austin P. An introduction to propensity score methods for reducing the effects of confounding in observational studies. Multivar Behav Res. 2011;46:399–424.
7. Rubin DB. The design versus the analysis of observational studies for causal effects: parallels with the design or randomized trials. Stat Med. 2007;26:20–36.
8. Cochran WG. The effectiveness of adjustment by subclassification in removing bias in observational studies. Biometrics. 1968;24:295–313.
9. Rosenbaum PR, Rubin DB. Reducing bias in observational studies using subclassification on the propensity score. J Am Stat Assoc. 1984;79:516–24.
10. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika. 1983;70:41–55.
11. D'Agostino RB Jr, D'Agostino RB. Estimating treatment effects using observational data. JAMA. 2007;297:314–6.
12. Little RJ, Rubin DB. Causal effects in clinical and epidemiological studies via potential outcomes: concepts and analytical approaches. Annu Rev Public Health. 2000;21:121–45.
13. Patrick AR, Schneeweiss S, Brookhart MA, Glynn RJ, Rothman KJ, Avorn J, et al. The implications of propensity score variable selection strategies in pharmacoepidemiology: an empirical illustration. Pharmacoepidemiol Drug Saf. 2011;20:551–9.
14. Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Sturmer T. Variable selection for propensity score models. Am J Epidemiol. 2006;163:1149–56.
15. Pattanayak CW, Rubin DB, Zell ER. Propensity score methods for creating covariate balance in observational studies. Rev Esp Cardiol. 2011;64:897–903.
16. Deb S, Austin PC, Tu JV, Ko DT, Mazer CD, Kiss A, et al. A review of propensity-score methods and their use in cardiovascular research. Can J Cardiol. 2016;32:259–65.
17. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med. 1996;15:361–87.
18. Cepeda MS, Boston R, Farrar JT, Strom BL. Comparison of logistic regression versus propensity score when the number of events is low and there are multiple confounders. Am J Epidemiol. 2003;158:280–7.
19. Muriel A, Penuelas O, Frutos-Vivar F, Arroliga AC, Abraira V, Thille AW, et al. Impact of sedative on outcomes of noninvasive ventilation. Intensive Care Med. 2015;41:1586–600.
20. Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology. 2000;11:550–60.
21. Delaney JW, Pinto R, Long J, Lamontagne F, Adhikari NK, Kumar A, et al. The influence of corticosteroid treatment on the outcome of influenza A (H1N1pdm09)-related critical illness. Crit Care. 2016;20:75.
22. Cole SR, Hernan MA. Constructing inverse probability weights for marginal structural models. Am J Epidemiol. 2008;168:656–64.
23. Morris JV, Ramnarayan P, Parslow RC, Fleming SJ. Outcomes for children receiving noninvasive ventilation as the first-line mode of mechanical ventilation at intensive care admission: a propensity score-matched cohort study. Crit Care Med. 2017;45:1045–53.
24. Rosa RG, Tonietto TF, Duso BA, Maccari JG, de Oliveira RP, Rutzen W, et al. Mortality of adult critically ill subjects with cancer. Respir Care. 2017;62:615–22.
25. Peñuelas O, Muriel A, Frutos-Vivar F, Fan E, Raymondos K, Rios F, et al. Prediction and outcome of intensive care unit-acquired paresis. J Intensive Care Med. 2018;33:16–28, http://dx.doi.org/10.1177/0885066616643529.

Chapter 3: Propensity Score compared to Randomized Control Trials to assess venous thromboembolism treatment on mortality

Chapter 3

In this chapter I present the study where we compared the results obtained when considering Propensity Score approach versus Randomized Control Trials. In particular we studied the effect of treatment for venous thromboembolism on patients' mortality.

3.1. Background

Venous thromboembolism (VTE) is a worldwide major health issue. Until the date, different clinical practice guidelines have been proposed in order to better manage it, but only few are supported by high-quality evidence studies. There is a need to further study potential guidelines to improve treatment decision for pulmonary embolism. RCTs are the gold standard to assess treatment effects.

3.2. Objectives

To assess differences in mortality outcomes reported by Propensity Score matching versus RCT using a systematic review and meta-analysis.

3.3. Methods

A research on PubMed and Web of Science was carried out to identify Propensity Score studies and RCT that were posteriorly matched based on the clinical treatment issue. Meta-analyses were run on these group of studies. Relative ORs were used as treatment efficacy measures to compare paired observational and RCT studies, where final summary relative ORs were also calculated.

3.4. Results

Seven clinical treatment issues were found and used to match observational and RCTs studies. Overall, we observed that Propensity Score and RCTs did not present differences in their mortality estimates [ROR=0.89; 95%CI (0.32;1.46)].

However, two of the seven treatment issues considered (thrombolysis vs anticoagulation for pulmonary embolism; once- vs twice-daily enoxaparin for VTE) showed differences in the treatment effect direction while two studies (rivaroxaban vs vitamin K antagonists for VTE; home treatment vs home hospitalization for DVT) showed non-significance difference in the magnitude of the effects.

3.5. Conclusions

This systematic comparison across seven VTE treatment topics suggests that the Propensity Score approach may lead to similar results in comparison to RCT. However, according to the treatment considered, differences in the direction of the magnitude of the treatment estimates may occur.

3.6. Publication

The full manuscript of our study is included in this chapter and published in the journal *Chest* (Impact Factor Journal Citation Reports 2019: 9.41; Q1; 8/139 Category *Pulmonary and Respiratory Medicine).*

Coscia C, Jaureguizar A, Quezada CA, Muriel A, Monreal M, Villén T, Barbero E, Chiluiza D, Yusen RD, Jimenez D. Comparison of All-Cause Mortality Following VTE Treatment Between Propensity Score-Adjusted Observational Studies and Matched Randomized Controlled Trials: Meta-Epidemiologic Study. Chest. 2019; 155(4):689-698. doi: 10.1016/j.chest.2018.10.016.

The citations counts of this article (21/11/21) provided by Google Scholar was 4 and by Scopus was 3.

CHEST

Comparison of All-Cause Mortality Following VTE Treatment Between Propensity Score-Adjusted Observational Studies and Matched Randomized Controlled Trials

Meta-Epidemiologic Study

Claudia Coscia, MD; Ana Jaureguizar, MD; Carlos Andres Quezada; Alfonso Muriel, PhD; Manuel Monreal, PhD; Tomas Villén, MD; Esther Barbero, MD; Diana Chiluiza, MD; Roger D. Yusen, MD; and David Jimenez, PhD

BACKGROUND: It is unknown whether propensity score-adjusted observational studies produce results comparable to those of randomized controlled trials (RCTs) that address similar VTE treatment issues.

METHODS: The PubMed and Web of Science databases were systematically searched for propensity score-adjusted observational studies, RCTs, and meta-analyses of RCTs that estimated all-cause mortality following VTE treatment. After identifying distinct clinical treatment issues evaluated in the eligible observational studies, a standardized algorithm was used to identify and match at least one RCT or RCT meta-analysis publication for paired study design analyses. Meta-analyses were used to summarize groups of studies. Treatment efficacy statistics (relative ORs) were compared between the paired observational and RCT studies, and the summary relative ORs for all study design pairs were also calculated.

RESULTS: The observational and RCT study pairs assessed seven clinical treatment issues. Overall, the observational study-RCT pairs did not exhibit significantly different mortality estimates (summary relative OR, 0.89; 95% CI, 0.32-1.46; $I^2 = 23\%$). However, two of the seven treatment issue study pairs (thrombolysis vs anticoagulation for pulmonary embolism; once- vs twice-daily enoxaparin for VTE) exhibited a significantly different treatment effect direction, and there was a substantial (nonsignificant) difference in the magnitude of the effect in another two of the study pairs (rivaroxaban vs vitamin K antagonists for VTE; home treatment vs hospitalization for DVT).

CONCLUSIONS: This systematic comparison across seven VTE treatment topics suggests that propensity score-adjusted observational studies and RCTs often exhibit similar all-cause mortality, although differences in the direction or the magnitude of estimated treatment effects may occasionally occur.

TRIAL REGISTRY: PROSPERO; CRD42018087819; URL: http://www.crd.york.ac.uk/PROSPERO.

CHEST 2019; 155(4):689-698

KEY WORDS: DVT; observational studies; pulmonary embolism; randomized controlled trials; VTE

ABBREVIATIONS: PE = pulmonary embolism; RCT = randomized controlled trial

AFFILIATIONS: From the Biostatistics Department (Drs Coscia and Muriel), Ramón y Cajal Hospital and Instituto Ramón y Cajal de Investigación Sanitaria IRYCIS, CIBERESP, Madrid, Spain; Respiratory Department (Drs Jaureguizar, Quezada, Barbero, Chiluiza, and Jimenez), Hospital Ramón y Cajal, Medicine Department, Universidad de Alcala (IRYCIS), Madrid, Spain; Department of Internal Medicine (Dr Monreal), Hospital Universitari Germans Trias i Pujol, Badalona, Barcelona, Universidad Católica de Murcia, Murcia, Spain; Emergency

Chapter 4: Mendelian Randomization and Causal Mediation Analysis

Chapter 4

In this chapter I present the study where we applied Mendelian Randomization and Causal Mediation Analysis to better study causal risk factors for pancreatic cancer, more specifically the role of type 2 diabetes mellitus subtypes and obesity.

4.1. Background

Pancreatic cancer (PC) presents a complex aetiology, with many potential risk factors involved in the development of the disease. Many efforts have been made in order to establish the relevant risk factors using observational studies but until the date only few studies applied causal inference to unravel these causal mechanisms. T2DM and obesity, are the most important risk factors for PC development and they present a complex relationship because they very often co-exist.

4.2. Objectives

The objective is to disentangle the causal effect of type 2 diabetes mellitus (T2DM) subtypes: long-standing (LSDM) and new onset (NODM) diabetes mellitus, on pancreatic cancer (PC), and to assess the potential role of BMI-obesity as mediator, applying causal

inference methodologies such as Mendelian Randomization (MR) and Causal Mediation Analysis (CMA).

4.3. Methods

Using the information from the PanGenEU European case-control study, with 2,018 cases and 1,540 controls, we applied the MR to study the causal effect of LSDM and NODM on PC, and vice versa, using genetic variants associated with T2DM-subtypes and PC. MR was also used to analyse the causality of obesity on PC. We considered MR-Egger regression, MR-IVW and multivariable MR to discard potential pleiotropy bias, as it was described in Chapter 1. Finally, we utilised Causal Mediation Analysis (CMA) to assess the potential mediator role of obesity between T2DM-subtypes and PC.

4.4. Results

Considering the results based on the MR technique, a causal effect of LSDM on PC could not be confirmed [OR=1.08; 95%CI (0.86;1.29)], however, we observed a significant causal effect of PC on NODM [OR=2.85, 95%CI (2.04;3.98)]. The same conclusion was obtained when multivariable MR was applied to discard potential pleiotropy bias [OR=1.31, 95%CI (1.10;1.52)].

Considering the CMA results, potential mediation effects of being overweight/obese were observed between NODM and PC [OR_{direct}=10.14; 95%CI (5.48;22.69); $OR_{indirect}$=0.55; 95%CI (0.23;0.92); OR_{total}=5.58; 95%CI (3.65;5.92)] and between LSDM and PC [OR_{direct} =1.61; 95%CI (1.31;2.00); $OR_{indirect}$=1.03; 95%CI (1.01;1.08); OR_{total}=1.67; 95%CI (1.35;2.06)].

4.5. Conclusions

Results based on the MR approach did not support a causal effect of LSDM on PC, however, our study showed that PC can cause NODM, suggesting that, in practice, NODM could be an early sign of PC. Moreover, being overweight/obese may be considered as a potential mediator between both T2DM-subtypes and PC. Nevertheless, more analyses should be carried out in order to further unravel these associations.

4.6. Publication

The full manuscript of our study is included in this chapter and was published in the journal *GUT* (Impact Factor Journal Citation Reports 2021: 23.059; Q1-first decile; 3/92 Category *Gastroenterology & Hepatology)*

Molina-Montes E, **Coscia C**, Gómez-Rubio P, Fernández A, Boenink R, Rava M, Márquez M, Molero X, Löhr M, Sharp L, Michalski CW, Farré A, Perea J, O'Rorke M, Greenhalf W, Iglesias M, Tardón A, Gress TM, Barberá VM, Crnogorac-Jurcevic T, Muñoz-Bellvís L, Dominguez-Muñoz JE, Renz H, Balcells J, Costello E, Ilzarbe L, Kleeff J, Kong B, Mora J, O'Driscoll D, Poves I, Scarpa A, Yu J, Hidalgo M, Lawlor RT, Ye W, Carrato A, Real FX, Malats N; PanGenEU Study Investigators. Deciphering the complex interplay between pancreatic cancer, diabetes mellitus subtypes and obesity/BMI through causal inference and mediation analyses. Gut. 2021 Feb;70(2):319-329. doi: 10.1136/gutjnl-2019-319990.

Despite being a recent publication, from February 2021, it is worth noting the high number of citations that it has received, 21 by Google Scholar and 13 by Scopus (21/11/21)

Supplementary Material of this manuscript is included in Appendix A.

ORIGINAL RESEARCH

Deciphering the complex interplay between pancreatic cancer, diabetes mellitus subtypes and obesity/BMI through causal inference and mediation analyses

Esther Molina-Montes,[1,2] Claudia Coscia,[1,2] Paulina Gómez-Rubio,[1,2] Alba Fernández,[1] Rianne Boenink,[1] Marta Rava,[1] Mirari Márquez,[1,2] Xavier Molero,[3,4] Matthias Löhr,[5] Linda Sharp ,[6,7] Christoph W Michalski,[8,9] Antoni Farré,[10] José Perea ,[11,12] Michael O'Rorke,[13,14] William Greenhalf,[15] Mar Iglesias,[2,16] Adonina Tardón,[17,18] Thomas M Gress,[19] Victor M Barberá,[20] Tatjana Crnogorac-Jurcevic,[21] Luis Muñoz-Bellvís,[22] J Enrique Dominguez-Muñoz,[23] Harald Renz,[24] Joaquim Balcells,[4,25] Eithne Costello ,[15] Lucas Ilzarbe,[26] Jörg Kleeff,[8,9] Bo Kong,[8] Josefina Mora,[10] Damian O'Driscoll,[6,7] Ignasi Poves,[27] Aldo Scarpa ,[28] Jingru Yu,[29] Manuel Hidalgo ,[30,31] Rita T Lawlor,[28] Weimin Ye,[29] Alfredo Carrato,[2,32] Francisco X Real,[2,33] Núria Malats ,[1,2] On behalf of the PanGenEU Study Investigators

► Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/gutjnl-2019-319990).

For numbered affiliations see end of article.

Correspondence to
Dr Núria Malats, Genetic and Molecular Epidemiology Group, Spanish National Cancer Research Center, Madrid E-28029, Spain; nuria@cnio.es

Received 3 October 2019
Revised 7 April 2020
Accepted 12 April 2020

To cite: Molina-Montes E, Coscia C, Gómez-Rubio P, *et al*. *Gut* Epub ahead of print: [*please include* Day Month Year]. doi:10.1136/gutjnl-2019-319990

ABSTRACT

Objectives To characterise the association between type 2 diabetes mellitus (T2DM) subtypes (new-onset T2DM (NODM) or long-standing T2DM (LSDM)) and pancreatic cancer (PC) risk, to explore the direction of causation through Mendelian randomisation (MR) analysis and to assess the mediation role of body mass index (BMI).

Design Information about T2DM and related factors was collected from 2018 PC cases and 1540 controls from the PanGenEU (European Study into Digestive Illnesses and Genetics) study. A subset of PC cases and controls had glycated haemoglobin, C-peptide and genotype data. Multivariate logistic regression models were applied to derive ORs and 95% CIs. T2DM and PC-related single nucleotide polymorphism (SNP) were used as instrumental variables (IVs) in bidirectional MR analysis to test for two-way causal associations between PC, NODM and LSDM. Indirect and direct effects of the BMI-T2DM-PC association were further explored using mediation analysis.

Results T2DM was associated with an increased PC risk when compared with non-T2DM (OR=2.50; 95% CI: 2.05 to 3.05), the risk being greater for NODM (OR=6.39; 95% CI: 4.18 to 9.78) and insulin users (OR=3.69; 95% CI: 2.80 to 4.86). The causal association between T2DM (57-SNP IV) and PC was not statistically significant (OR_{LSDM}=1.08, 95% CI: 0.86 to 1.29, OR_{NODM}=1.06, 95% CI: 0.95 to 1.17). In contrast, there was a causal association between PC (40-SNP IV) and NODM (OR=2.85; 95% CI: 2.04 to 3.98), although genetic pleiotropy was present (MR-Egger: p value=0.03). Potential mediating effects of BMI (125-SNPs as IV), particularly in terms of weight loss, were evidenced on the NODM-PC association (indirect effect for BMI in previous years=0.55).

Conclusion Findings of this study do not support a causal effect of LSDM on PC, but suggest that PC causes NODM. The interplay between obesity, PC and T2DM is complex.

Significance of this study

What is already known on this subject?

- The association between type 2 diabetes mellitus (T2DM) and risk of pancreatic cancer (PC) has been evidenced in numerous studies.
- Few studies have suggested the distinct role of T2DM subtypes, new-onset T2DM (NODM) and long-standing T2DM (LSDM), in PC aetiology; while both were associated with PC risk, they may exhibit a different causal relationship with PC.
- Uncertainties surrounding the association between T2DM and PC risk also concern confounding or mediation by obesity, T2DM medication effects and the causal pathway linking both diseases.

INTRODUCTION

Pancreatic cancer (PC) has a high case-fatality rate in Western countries,[1] expected to rise in coming years if no immediate actions are taken.[2 3] Many unknowns in PC aetiology remain even regarding some of the well-established risk factors of this disease.[4] This also applies to type 2 diabetes mellitus (T2DM), despite representing an

Chapter 5: Mendelian Randomization in Causal Mediation Analysis, ***MRinCMA***

Chapter 5

In this chapter I present a methodological study where a new approach was proposed in order to answer mediation analysis questions. This approach consists of using genetic variants as instruments, like in the Mendelian Randomization, to study the potential mediation effects. We run several simulation studies to assess our proposal and we applied it to study the causal effect of BMI/obesity on pancreatic cancer considering LSDM as mediator.

5.1. Abstract

Background

Mendelian Randomization (MR) is the use of genetic variants as instrumental variables to evaluate the causal effect of an exposure on an outcome. On the other hand, Causal Mediation Analysis (CMA) is used to study the causal effect of the exposure on the outcome, considering a third variable, the mediator, that occurs in between. In CMA three different effects can be estimated: direct, indirect, and total effect. However, strong assumptions surround CMA. Further developments of CMA are needed in order to obtain unbiased effects.

Objectives

The use of Causal Mediation Analysis (CMA) to obtain the direct, indirect, and total effects of an exposure on an outcome, considering the mediation effect of a third variable, is increasing in epidemiological studies but it requires strong assumptions on confounding

bias. The aim of this study is to propose a new alternative as an extension of CMA and Mendelian Randomization (MR), that uses genetics variants as instrumental variables, and to analyse the causal effect of BMI/obesity or diabetes on pancreatic cancer, considering each factor as potential mediators.

Methods

We propose the *MRinCMA* approach to obtain unbiased estimates of the direct, indirect, and total effects. This proposal consists of using two genetic instruments, for exposure and mediator respectively, relaxing the CMA assumptions. We generate several simulated datasets based on the PanGenEU study features, modifying the type of variables, either continuous or categorical, correlations, sample sizes and effect sizes. We compare the results obtained using this new approach with the Structural Equation Model (SEM). We assess the proposed method's performance calculating values of mean squared error, coverage, bias-eliminated coverage, and empirical standard errors, as performance measures. We apply *MRinCMA* to analyse the effect of BMI/obesity on PC, with the mediation effect of LSDM.

Results

In general, unbiased estimates for direct, indirect, and total effect were obtained using *MRinCMA*. When continuous variables were considered, *MRinCMA* and SEM produce similar values across the performance measures, suggesting that both approaches are valid to obtain unbiased estimates. When at least one non-continuous variables was considered, MRinCMA presented overall lower bias and better coverage values than SEM. When analysing the effect of obesity on PC considering LSDM, as a potential mediator, we did not find any evidence of causality [OR_{direct}=0.97; 95%CI (0.62; 1.53); $OR_{indirect}$=1.13; 95%CI (0.89;1.88); OR_{total}=1.12; 95%CI (0.72; 1.65)].

Conclusions

With this new approach, *MRinCMA*, we showed that researchers may obtain unbiased estimates for the direct, indirect, and total effect using two independent genetic instruments, by appropriately accounting for the confounding bias assumption under different scenarios.

Publication

This study has been submitted to American Journal of Epidemiology.

5.2 Introduction

In epidemiological studies, an important goal is to analyse the causal relationship between an exposure and an outcome. In general, epidemiological analysis relies on observational data and may present bias generated by other factors that are associated with the outcome and the exposure and can distort the real effect of the exposure (Hernán and Robins 2020; Imbens and Rubin 2015; Pearl 2010a). According to the availability of the information on these uncontrolled factors, they are defined as measured (VanderWeele 2016) (i.e., all the confounders are collected and measured) or unmeasured (Hernán and Robins 2020; Imbens and Rubin 2015; Pearl 2010a). Several statistical methods have been developed for dealing with intermediate and confounder variables, among them Causal Mediation Analysis for measured confounding, and Mendelian Randomization and Structural Equation Models (Belope 2019; Muthén and Asparouhov 2015; Pearl 2012a) for unmeasured confounding.

As it was already explained in Chapter 1, the Causal Mediation Analysis (CMA) is used to estimate the causal effect of the exposure on the outcome considering the effect of a third variable, called mediator, which occurs in the pathway from the exposure to the outcome (Valeri and VanderWeele 2013; VanderWeele 2016; VanderWeele and Vansteelandt 2009). CMA estimates three parameters: a) the direct effect: effect of the exposure on the outcome independent of the mediator, b) the indirect effect: effect of the exposure on the outcome that occurs through the mediator, and c) the total effect: effect of the exposure on the outcome considering all previous effects (Figure 12A). They usually are estimated by either the Product or the Difference methods proposed by Baron and Kenny (BK) (De Stavola et al. 2015; VanderWeele and Vansteelandt 2009) according to the nature of the outcome. Strong assumptions are required when applying CMA (i.e., all the confounding variables are known and measured), which are difficult to meet (Carter et al. 2021) and may lead to biased and not representative results.

Mendelian Randomization (MR) analysis is extensively used in epidemiology to analyse causality between an exposure and an outcome (Burgess et al. 2013; Davies et al. 2018). This procedure applies an Instrumental Variable (*IV*) approach leading to unbiased estimates of the causal effect, even when there is unmeasured confounding bias (S Burgess et al. 2017; Burgess and Thompson 2013; Davies et al. 2018; Uddin et al. 2015). The MR uses either genetic variants or a genetic score (Burgess and Thompson 2013), as *IV* to estimate the causal effect of the exposure on an outcome (Figure 12B), assuming that: i) the *IV* is associated with the exposure, ii) the *IV* is associated with the outcome only through the exposure, and iii) the *IV* is not associated with any confounder (Burgess and Thompson 2013). One of the classical estimation methods applied in MR is the 2-Stage-Least-Squares (2SLS). The 2SLS is a regression-based method consisting in using the

predicted values of the exposure to estimate the causal effect on the outcome (S Burgess et al. 2017; Uddin et al. 2015). There are some extensions of the 2SLS depending on the type of the variables, such as the 2-Stage-Residual-Inclusion (2SRI), which is a good alternative for non-continuous variables (Palmer et al. 2017; Terza et al. 2008).

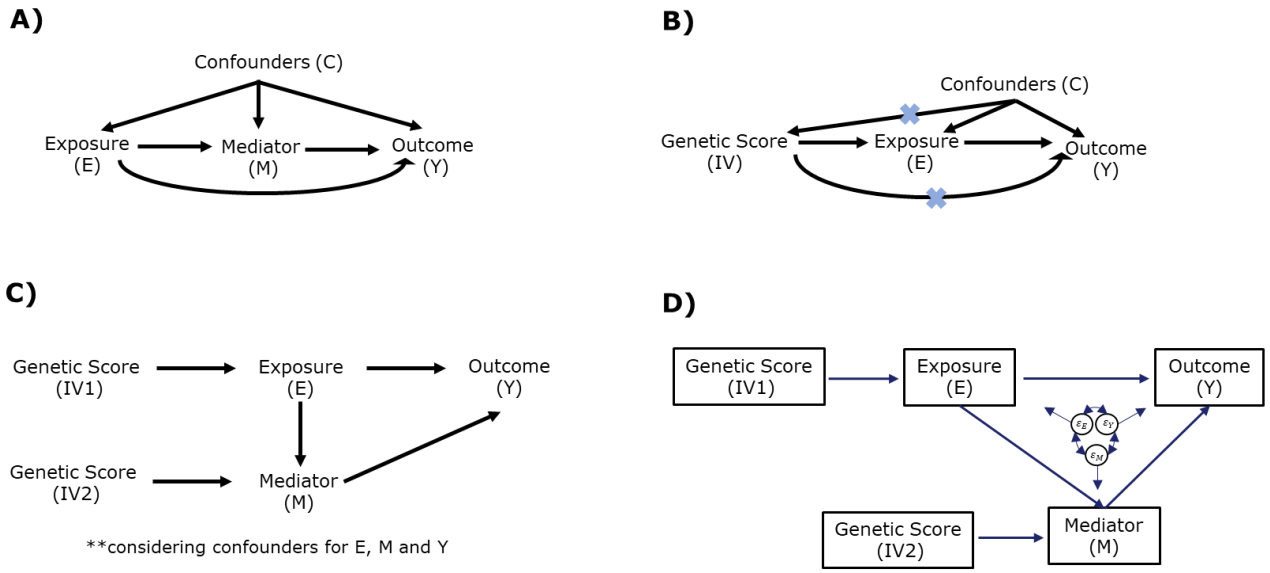


Figure 12. Causal diagrams
A) Causal Mediation Analysis (*CMA*) graph. B) Mendelian Randomization graph. C) Mediation Analysis with Instrumental Variables IVs (*MRinCMA* as specific case). D) Structural Equation Model *SEM*, where directed arrows represent causal effect, bidirectional arrows indicate a correlation between variables. Measurement variables are represented by squares and latent variables i.e., not observed are represented by circles.

Some authors already suggested that using *IV*s in the CMA context could be a good alternative to relax the CMA assumptions and to obtain unbiased results (Carter et al. 2021). Burgess et al. (2015), (Figure 12C), proposed an extension of the 2SLS regression-based method adding a third stage (i.e., 3SLS) when regressing continuous outcome variables, and they compared the performance of this method to SEM (Figure 12D) when analysing the causality between Body Mass Index (BMI), C-reactive protein, and uric acid. Also, in the context of continuous outcomes. Frölich and Huber (2017) applied a four-stage extension (i.e., 4SLS and 4SRI) to analyse the effect of education and income on the social functioning using school leaving and windfall income as IVs. While Burgess et al. (2015), considered the exposure, mediator and outcome as continuous variables, Frölich and Huber (2017) included the exposure as binary, maintaining mediator and outcome as continuous. Furthermore, North et al. (2019) used the same causal diagram as Burgess et al. (2015) to estimate the interaction terms between the exposure and the mediator considering two different instruments. Other authors like Sanderson (2021) proposed an approach based on multivariable MR (MVMR), while Relton et al. (2012) considered a two-step MR. Finally,

we made a comprehensive summary of the current methods using MR in mediation analysis and they proposed an alternative combining both approaches.

Up to now, only Carter et al. (2021) have suggested an alternative for categorical outcomes, indicating a lack of development of methods when modelling categorical variables, despite the fact that such variables are those commonly used in epidemiological studies. For that reason, in this paper, we proposed to extend Burgess et al. (2015) and Frölich and Huber (2017) approaches, offering a simple, valid and flexible methodology that would consider not only continuous and normally distributed variables, but also categorical variables.

Therefore, the aim of this study was to propose four extensions of the CMA, which we named "Mendelian Randomization in Causal Mediation Analysis" (*MRinCMA*), by applying MR to correct for potential confounder bias. We estimated direct, indirect, and total effects by incorporating two *IV*s in the CMA context, considering as *MRinCMA* approaches the extension of the 2SLS (3SLS and 4SLS) and 2SRI (3SRI and 4SRI) methods. We compared their performances to SEM, as a reference method. Furthermore, we applied *MRinCMA* to assess causality between BMI/obesity, type-2 diabetes mellitus (T2DM), and pancreatic cancer (PC) risk in the context of a previously published CMA and MR effort conducted within the PanGenEU case-control study (Molina-Montes et al. 2021).

5.3. Methods

5.3.1. The *MRinCMA* approaches to estimate direct, indirect, and total effects

We define four *MRinCMA* approaches: the 4-Stage-Least-Squares (4SLS), 3-Stage-Least-Squares (3SLS), 4-Stage-Residual-Inclusion (4SRI), and 3-Stage-Residual-Inclusion (3SRI). The causal graph is presented in Figure 12C, where Y is the outcome; E, the exposure; M, the mediator; IV_1, the exposure weighted genetic score; IV_2, the mediator weighted genetic score; and C, the vector of confounders.

The 4SLS approach required fitting the following four regression models to estimate direct, indirect, and total effects:

1. First model:

$$E = \beta_0 + \beta_1 IV_1 + \beta_2 C + \varepsilon_1 \quad [30]$$

2. Second model:

$$M = \mu_0 + \mu_1 \hat{E} + \mu_2 IV_2 + \mu_3 C + \varepsilon_2 \quad [31]$$

3. Third model:

$$Y = \alpha_0 + \alpha_1\hat{E} + \alpha_2\hat{M} + \alpha_3 C + \varepsilon_3 \quad [32]$$

4. Fourth model:

$$M = \gamma_0 + \gamma_1\hat{E} + \gamma_2 C + \varepsilon_4 \quad [33]$$

where $\hat{E}$ and $\hat{M}$ are the predictor vectors obtained from [30] and [31] respectively.

Based on BK methodology (De Stavola et al. 2015; VanderWeele and Vansteelandt 2009), the direct effect is the estimated effect of $\hat{E}$ over Y, $\hat{\alpha}_1$, derived from [32]; the indirect effect is the product $\left(\hat{\gamma}_1 \times \hat{\alpha}_2\right)$, where $\hat{\gamma}_1$ is the estimated effect of $\hat{E}$ over M, derived from [33]; and $\hat{\alpha}_2$ is the estimated effect of $\hat{M}$ over Y, derived from [32]. The total effect is the sum of both effects, $\hat{\alpha}_1 + \left(\hat{\gamma}_1 \times \hat{\alpha}_2\right)$ for continuous outcomes and the product $\hat{\alpha}_1 \times \left(\hat{\gamma}_1 \times \hat{\alpha}_2\right)$ for binary outcomes.

The 3SLS approach only needs three regression models, [30] to [32]. The direct effect is $\hat{\alpha}_1$ the indirect effect is $\left(\hat{\mu}_1 \times \hat{\alpha}_2\right)$, and the total effect is the sum, or the product, as appropriate, of both effects.

The 4SRI approach is similar to 4SLS but replacing the fitted values $(\hat{E}, \hat{M})$ by both the observed values (E, M) and the residuals $\left(\hat{\varepsilon}_1, \hat{\varepsilon}_2\right)$, therefore fitting the following models:

$$E = \tau_0 + \tau_1 IV_1 + \tau_2 C + \varepsilon_1 \quad [34]$$

$$M = \eta_0 + \eta_1 E + \eta_2\hat{\varepsilon}_1 + \eta_3 IV_2 + \eta_4 C + \varepsilon_2 \quad [35]$$

$$Y = \nu_0 + \nu_1 E + \nu_2\hat{\varepsilon}_1 + \nu_3 M + \nu_4\hat{\varepsilon}_2 + \nu_5 C + \varepsilon_3 \quad [36]$$

$$M = \lambda_0 + \lambda_1 E + \lambda_2\hat{\varepsilon}_1 + \lambda_3 C + \varepsilon_4 \quad [37]$$

where $\hat{\varepsilon}_1$ and $\hat{\varepsilon}_2$ are the residual vectors obtained from [34] and [35] respectively.

The direct effect is $\hat{\nu}_1$; the indirect effect is $\left(\hat{\lambda}_1 \times \hat{\nu}_3\right)$ and the total effect is the sum of both, $\hat{\nu}_1 + \left(\hat{\lambda}_1 \times \hat{\nu}_3\right)$ or the product $\hat{\nu}_1 \times \left(\hat{\lambda}_1 \times \hat{\nu}_3\right)$ for continuous and categorical outcomes, respectively.

Analogously to 3SLS, the 3SRI method only requires models [34] to [36] The direct effect is $\hat{\nu}_1$, the indirect effect is $\left(\hat{\eta}_1 \times \hat{\nu}_3\right)$, and the total effect is the sum, or the product, as appropriate, of both effects.

5.3.2. Structural Equation Model as a reference method

We also estimated the three effects using the Structural Equation Model (SEM) as it was previously done by Burgess et al. (2015). It is a multivariate technique widely used (Belope 2019; Pearl 2012a), that specifies measurement errors of the variables and permits to incorporate both observed and unobserved variables (Muthén 1984, 2011; Pearl 2010a, 2012a).

We defined the equations as shown in Figure 12D. Coefficient estimates of E and M in SEM represent direct and indirect effects; and the total effect is the sum of both assuming no interaction. The Maximum Likelihood (ML) approximation is used for continuous and normally-distributed variables while the Weighted Least Squares (WLS) method was considered for categorical variables (Li 2016; Muthén 1984; Muthén and Asparouhov 2015; Olsson et al. 2000).

5.3.3. Simulation efforts

Simulation studies are an important tool for assessing the performance of statistical methods (Morris et al. 2019). We evaluated the performance of the five aforementioned methods in estimating our parameters of interest: direct, indirect, and total effects, under different scenarios. Summary is shown in Figure 13.

Datasets were generated based on the epidemiological, clinical, and genetic data gathered from the PanGenEU study, a European case-control study including 2500 PC patients and 1500 controls (López de Maturana et al. 2021; Molina-Montes et al. 2021) (Appendix B).

Independent variables and outcomes of interest. The outcomes, exposures, and mediators have been defined as either quantitative or binary variables. We considered two outcomes, a binary trait reflecting the PC status and a continuous one corresponding to the pancreatic cancer risk score. BMI (continuous) and obesity (categorized as 1 if BMI >30 and 0 otherwise (Kent et al. 2017; Molina-Montes et al. 2021)) were the exposures (E). Long-standing diabetes mellitus (LSDM) (i.e., diabetes diagnosed > 2 years before the study recruitment (Molina-Montes et al. 2021) and categorized as yes/no) and glycated haemoglobin (HbA1c, continuous) were considered as mediators (M). We included two continuous genetic scores (IV_1 and IV_2) weighting by the effect of each genetic variant on

the exposure and the mediator respectively. We incorporated sex (categorical: man/woman), age (continuous), and smoking status (categorical: non-smoker/occasional/former/ current smoker) as confounding variables (C).

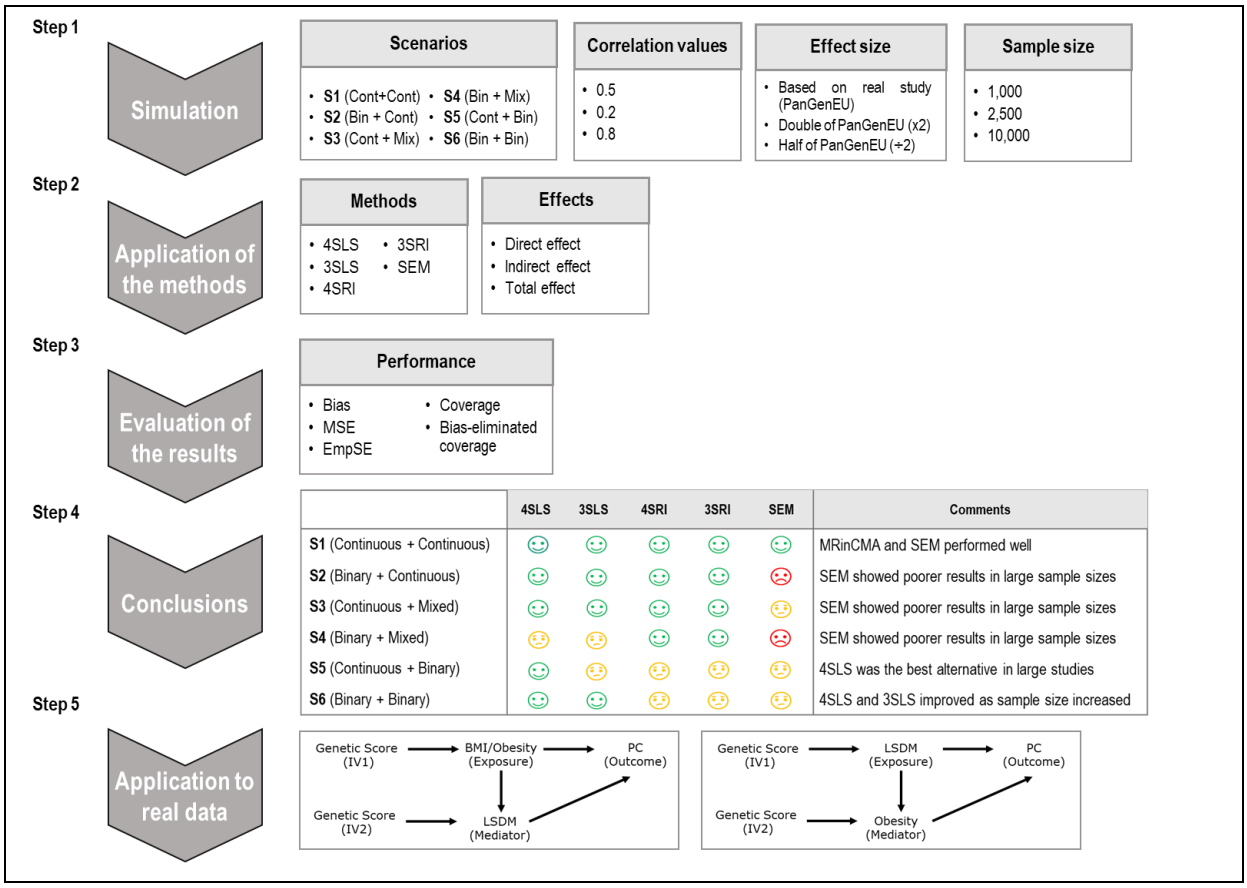


Figure 13. Simulations steps and workflow.
S1-S6: Different simulation scenarios; 4SLS=4-Stage-Least-Squares; 3SLS=3-Stage-Least-Squares; 4SRI=4-Stage-Residual-Inclusion; 3SRI=3-Stage-Residual-Inclusion; SEM= Structural Equation Models; MSE=Mean Squared Error; EmpSE: Empirical standard error; BMI: Body Mass Index; LSDM: Long-standing Diabetes Mellitus; PC: Pancreatic Cancer

Data-generating mechanisms. According to the definition of the variables $Y, E,$ and M, the following six scenarios were studied:

S1: Y, E, and M as continuous.

S2: Y as binary and both E and M as continuous.

S3: Y as continuous, E as continuous, and M as binary.

S4: Y as binary, E as continuous, and M as binary.

S5: Y as continuous and both E and M as binary.

S6: Y, E, and M as binary.

Variables $Y, E,$ and M were simulated as continuous in all cases, using probit and logit transformation (Y*, E* and M*) for the binary settings. Therefore, initial linear regression models were assumed regardless of the character of these variables. The original Y, E and M were recovered subsequently by inverting this transformation (either probit or logit), obtaining the final binary variables of interest. Note that Y equalled Y*, E equalled E*, and M* equalled M when they were continuous.

We simulated Y*, E* and M* using a multivariate normal model, assuming random errors with vectors of means $\mu = (0,0,0)$ and variance-covariance matrix:

$$\Sigma = \begin{pmatrix} \sigma_E^2 & \rho\sigma_E\sigma_M & \rho\sigma_E\sigma_Y \\ \rho\sigma_E\sigma_M & \sigma_M^2 & \rho\sigma_M\sigma_Y \\ \rho\sigma_E\sigma_Y & \rho\sigma_M\sigma_Y & \sigma_Y^2 \end{pmatrix}$$

Standard deviations $(\sigma_E, \sigma_M, \sigma_Y)$ were considered, based on PanGenEU results. Correlation, ρ, ranged from weak (0.2), and moderate (0.5) to strong (0.8).

We proposed different simulation settings according to: i) the nature of the variables (S1 to S6), ii) the correlation between variables $\rho = (0.2, 0.5, 0.8)$, iii) the effect sizes (based on PanGenEU estimates, half and double of PanGenEU estimates), and iv) the study sample size ($1{,}000$, $2{,}500$ and $10{,}000$). Each setting was repeated m=$2{,}000$ times. Additional details about the simulation procedure are available in Appendix C and values of the prefixed coefficients considered are shown in Table D1 (Appendix D). When dealing with categorical variables, the probit transformation was always considered for SEM, while for *MRinCMA*, we considered the probit transformation for scenarios S3 and S5 (continuous outcome) and logit transformation otherwise, because better performance results were observed.

5.4. Performance indicators

We evaluated the five approaches using the following metrics: bias, mean squared error (MSE), empirical standard error (EmpSE), 95% confidence interval coverage rate (CI-C), and bias-eliminated coverage rate (BE-C). A detailed definition of these performance measures is presented in Appendix E and in Morris et al. (2019).

All analyses were performed using R software version 4.0.1. SEM were fitted using package *lavaan* (Rosseel 2012) which implements WLS for categorical data fitting a probit regression model.

5.5. Results

5.5.1. Results from the simulation study

Here we present a summary of the performance of the five methods, also described in Figure 13.

S1: Y, E, and M as continuous. All methods yielded a null or negligible bias in the estimation of the three effects, Table 4. In Tables 5 and Table D2 we present the estimated 95% confidence interval coverage rate (CI-C) and bias-eliminated coverage rate (BE-C). For the direct effect, SEM outperformed the *MRinCMA* methods with results around the nominal 95%CI cut-off point. On the other hand, for the indirect and total effects, we noted that CI-C and BE-C were reasonably close to the cut-off point for the five methods.

S2: Y as binary and both E and M as continuous. The results obtained for *MRinCMA* methods were similar to those from *S1*, with good performance results overall, Tables 5, 6 and Table D2. By contrary, SEM showed higher bias and presented poor results with CI-C deteriorated to a clearly unacceptable coverage for indirect and total effects when the sample size increased, Tables 5 and Table D2.

S3: Y and E as continuous, and M as binary. All methods provided unbiased estimates, Tables 7 and Table D3. The coverage rates for *MRinCMA* approaches were better, SEM led to rates utterly different from the nominal level (95%), regardless of the sample size, Tables 5 and Table D2.

S4: Y as binary, E as continuous, and M as binary. In this case, 4SRI and 3SRI methods achieved the best results for direct and indirect effects in terms of bias, Tables 7 and Table D3. For the total effect, results obtained with SEM were better, but this method presented poor coverage rates, Table 5, worsening with large sample size, i.e., n=10,000, Table D2.

S5: Y as continuous, and E and M as binary. Biased estimates and high MSE values were obtained in both *MRinCMA* and SEM, Table 8. SEM showed poorer coverage rates than *MRinCMA,* Table 4. 4SLS was the best method with large sample size, Table D2 and D3.

*S6: Y, E, and M as binary. S*imilar conclusions as those obtained in scenario S5 could be derived here, 4SLS was, in general, the best option, Tables 5 and 8. Its performance improved as sample size increased, Tables D2 and D3.

A sensitivity analysis has been done considering higher and lower effect sizes and stronger or weaker correlations between Y, E, and M, and we obtained similar performance results (data not shown).

	Direct			Indirect			Total		
	Bias	**MSE**	**EmpSE**	**Bias**	**MSE**	**EmpSE**	**Bias**	**MSE**	**EmpSE**
n = 1,000	**Estimate (SE)**	**Estimate (SE)**	**Estimate (SE)**	**Estimate (SE)**	**Estimate (SE)**	**Estimate (SE)**	**Estimate (SE)**	**Estimate (SE)**	**Estimate (SE)**
4SLS	-0.001 (0.001)	0.003 (0.000)	0.055 (0.001)	0.002 (0.001)	0.001 (0.000)	0.030 (0.000)	0.001 (0.001)	0.004 (0.000)	0.066 (0.001)
3SLS	-0.001 (0.000)	0.003 (0.000)	0.055 (0.001)	-0.004 (0.001)	0.001 (0.000)	0.029 (0.000)	-0.004 (0.001)	0.004 (0.000)	0.066 (0.001)
4SRI	-0.001 (0.001)	0.003 (0.000)	0.053 (0.001)	0.002 (0.001)	0.001 (0.000)	0.028 (0.000)	0.001 (0.001)	0.004 (0.000)	0.066 (0.001)
3SRI	-0.001 (0.001)	0.003 (0.000)	0.053 (0.001)	-0.003 (0.001)	0.001 (0.000)	0.027 (0.000)	-0.004 (0.001)	0.004 (0.000)	0.066 (0.001)
SEM	-0.001 (0.001)	0.003 (0.000)	0.053 (0.001)	-0.003 (0.001)	0.001 (0.000)	0.027 (0.000)	-0.004 (0.001)	0.004 (0.000)	0.066 (0.001)
n = 2,500									
4SLS	0.000 (0.001)	0.001 (0.000)	0.034 (0.000)	0.004 (0.000)	0.000 (0.000)	0.018 (0.000)	0.004 (0.001)	0.002 (0.000)	0.040 (0.001)
3SLS	0.000 (0.001)	0.001 (0.000)	0.034 (0.000)	-0.001 (0.000)	0.000 (0.000)	0.017 (0.000)	-0.002 (0.001)	0.002 (0.000)	0.040 (0.001)
4SRI	-0.001 (0.001)	0.001 (0.000)	0.033 (0.000)	0.004 (0.000)	0.000 (0.000)	0.017 (0.000)	0.004 (0.001)	0.002 (0.000)	0.040 (0.001)
3SRI	-0.001 (0.000)	0.001 (0.000)	0.033 (0.000)	-0.001 (0.000)	0.000 (0.000)	0.016 (0.000)	-0.002 (0.001)	0.002 (0.000)	0.040 (0.001)
SEM	-0.001 (0.001)	0.001 (0.000)	0.033 (0.000)	-0.001 (0.000)	0.000 (0.000)	0.016 (0.000)	-0.002 (0.001)	0.002 (0.000)	0.040 (0.001)
n = 10,000									
4SLS	0.000 (0.000)	0.000 (0.000)	0.016 (0.000)	0.005 (0.000)	0.000 (0.000)	0.009 (0.000)	0.005 (0.000)	0.000 (0.000)	0.019 (0.000)
3SLS	0.000 (0.000)	0.000 (0.000)	0.016 (0.000)	0.000 (0.000)	0.000 (0.000)	0.008 (0.000)	0.000 (0.000)	0.000 (0.000)	0.019 (0.000)
4SRI	0.000 (0.000)	0.000 (0.000)	0.016 (0.000)	0.005 (0.000)	0.000 (0.000)	0.008 (0.000)	0.005 (0.000)	0.000 (0.000)	0.019 (0.000)
3SRI	0.000 (0.000)	0.000 (0.000)	0.016 (0.000)	0.000 (0.000)	0.000 (0.000)	0.008 (0.000)	0.000 (0.000)	0.000 (0.000)	0.019 (0.000)
SEM	0.000 (0.000)	0.000 (0.000)	0.016 (0.000)	0.000 (0.000)	0.000 (0.000)	0.008 (0.000)	0.000 (0.000)	0.000 (0.000)	0.019 (0.000)

Table 4. Bias, Mean Squared Error (MSE), and Empirical Standard Error (EmpSE) in Scenario S1 according to different sample sizes, based on coefficients shown in Table D1. S1: *Y*, *E* and *M* continuous; MSE=Mean Squared Error; EmpSE=Empirical Standard Error; SE=Standard Error; 4SLS=4-Stage-Least-Squares; 3SLS=3-Stage-Least-Squares; 4SRI=4-Stage-Residual-Inclusion; 3SRI=3-Stage-Residual-Inclusion; SEM= Structural Equation Models.

S1	n = 1,000						n= 2,500					
	Direct		Indirect		Total		Direct		Indirect		Total	
	CI-C	BE-C	CI-C	BE-C	CI-C	BE-C	CI-C	BE-C	CI-C	BE-C	CI-C	BE-C
4SLS	99.40%	99.40%	93.70%	93.80%	92.00%	92.10%	99.60%	99.60%	93.80%	93.00%	91.90%	92.55%
3SLS	99.40%	99.40%	92.20%	93.60%	91.50%	91.75%	99.60%	99.60%	92.70%	93.25%	92.90%	92.70%
4SRI	88.30%	88.65%	93.30%	92.70%	92.60%	92.30%	88.30%	87.95%	93.00%	93.20%	91.90%	92.40%
3SRI	88.30%	88.65%	91.40%	92.65%	92.50%	92.50%	88.30%	87.95%	92.60%	92.85%	92.50%	92.45%
SEM	94.70%	94.78%	91.40%	92.00%	92.10%	91.95%	94.70%	94.40%	93.10%	93.45%	93.10%	93.10%
S2	n = 1,000						n= 2,500					
	Direct		Indirect		Total		Direct		Indirect		Total	
	CI-C	BE-C	CI-C	BE-C	CI-C	BE-C	CI-C	BE-C	CI-C	BE-C	CI-C	BE-C
4SLS[a]	99.00%	99.00%	93.00%	93.00%	92.00%	92.00%	99.00%	99.00%	94.00%	93.00%	91.00%	92.00%
3SLS[a]	99.00%	99.00%	92.00%	93.00%	92.00%	92.00%	99.00%	99.00%	94.00%	93.00%	93.00%	93.00%
4SRI[a]	88.00%	89.00%	94.00%	94.00%	92.00%	92.00%	88.00%	88.00%	93.00%	93.00%	93.00%	92.00%
3SRI[a]	88.00%	89.00%	92.00%	93.00%	92.00%	92.00%	88.00%	88.00%	93.00%	93.00%	93.00%	92.00%
SEM[b]	96.00%	97.00%	78.00%	94.00%	96.00%	99.00%	94.00%	95.00%	67.00%	92.00%	78.00%	97.00%
S3	n = 1,000						n= 2,500					
	Direct		Indirect		Total		Direct		Indirect		Total	
	CI-C	BE-C	CI-C	BE-C	CI-C	BE-C	CI-C	BE-C	CI-C	BE-C	CI-C	BE-C
4SLS[b]	96.90%	97.00%	98.80%	98.80%	96.50%	97.00%	97.20%	97.20%	92.10%	95.70%	92.90%	94.70%
3SLS[b]	96.90%	97.00%	99.00%	99.10%	95.60%	97.00%	97.20%	97.20%	92.00%	96.60%	92.30%	94.90%
4SRI[b]	90.70%	90.50%	95.00%	94.30%	93.30%	93.40%	91.90%	91.90%	93.20%	92.90%	92.80%	93.10%
3SRI[b]	90.70%	90.50%	93.80%	93.80%	93.20%	93.30%	91.90%	91.90%	93.30%	92.80%	93.40%	93.70%
SEM[b]	97.00%	95.20%	79.80%	94.00%	86.60%	91.70%	96.10%	96.10%	38.50%	96.70%	76.70%	93.30%

S4	n = 1,000						n= 2,500					
	Direct		Indirect		Total		Direct		Indirect		Total	
	CI-C	BE-C	CI-C	BE-C	CI-C	BE-C	CI-C	BE-C	CI-C	BE-C	CI-C	BE-C
4SLS[a]	95.00%	95.00%	98.00%	98.00%	98.00%	98.00%	94.00%	94.00%	93.00%	96.00%	89.00%	96.00%
3SLS[a]	95.00%	95.00%	99.00%	99.00%	97.00%	98.00%	94.00%	94.00%	94.00%	96.00%	90.00%	97.00%
4SRI[a]	95.00%	95.00%	98.00%	98.00%	93.00%	96.00%	95.00%	95.00%	94.00%	94.00%	82.00%	96.00%
3SRI[a]	95.00%	95.00%	99.00%	99.00%	94.00%	96.00%	95.00%	95.00%	93.00%	95.00%	83.00%	96.00%
SEM[b]	96.00%	96.00%	79.00%	99.00%	93.00%	100.00%	96.00%	97.00%	36.00%	98.00%	51.00%	99.00%
S5	**n = 1,000**						**n= 2,500**					
	Direct		Indirect		Total		Direct		Indirect		Total	
	CI-C	BE-C	CI-C	BE-C	CI-C	BE-C	CI-C	BE-C	CI-C	BE-C	CI-C	BE-C
4SLS[b]	97.20%	96.90%	99.50%	99.50%	98.70%	98.30%	97.10%	96.20%	98.80%	98.80%	95.90%	96.30%
3SLS[b]	97.20%	96.90%	99.50%	99.50%	98.10%	98.00%	97.10%	96.20%	98.80%	98.70%	95.90%	95.60%
4SRI[b]	94.40%	94.10%	99.60%	99.60%	98.80%	98.70%	93.10%	94.00%	99.40%	99.50%	94.80%	96.60%
3SRI[b]	94.40%	94.10%	99.70%	99.80%	98.40%	99.00%	93.10%	94.00%	99.60%	99.60%	95.00%	96.60%
SEM[b]	91.90%	91.60%	73.70%	98.60%	93.60%	95.20%	93.60%	94.40%	39.30%	96.00%	81.80%	93.40%
S6	**n = 1,000**						**n= 2,500**					
	Direct		Indirect		Total		Direct		Indirect		Total	
	CI-C	BE-C	CI-C	BE-C	CI-C	BE-C	CI-C	BE-C	CI-C	BE-C	CI-C	BE-C
4SLS[a]	95.00%	95.00%	99.00%	99.00%	98.00%	97.00%	95.00%	95.00%	99.00%	99.00%	95.00%	96.00%
3SLS[a]	95.00%	95.00%	99.00%	99.00%	98.00%	98.00%	95.00%	95.00%	99.00%	99.00%	97.00%	97.00%
4SRI[a]	95.00%	95.00%	99.00%	99.00%	99.00%	99.00%	96.00%	96.00%	99.00%	99.00%	97.00%	98.00%
3SRI[a]	95.00%	95.00%	99.00%	99.00%	99.00%	99.00%	96.00%	96.00%	99.00%	99.00%	97.00%	98.00%
SEM[b]	99.00%	99.00%	89.00%	99.00%	97.00%	100.00%	96.00%	98.00%	57.00%	99.00%	75.00%	99.00%

Table 5. 95% confidence interval coverage rate and bias-eliminated coverage for the direct, indirect, and total effect estimation in Scenarios S1-S6 and according to different sample sizes, based on PanGenEU coefficients shown in Table D1.
S1: *Y*, *E* and M continuous; S2: *Y* binary and *E* and *M* continuous; S3: *Y* continuous, *E* continuous and *M* binary; S4: *Y* binary, *E* continuous and *M* binary; S5: *Y* continuous and *E* and *M* binary; S6: *Y*, *E* and *M* binary; CI-C= 95% confidence interval coverage rate; BE-C= Bias-Eliminated coverage; 4SLS=4-Stage-Least-Squares; 3SLS=3-Stage-Least-Squares; 4SRI=4-Stage-Residual-Inclusion; 3SRI=3-Stage-Residual-Inclusion; SEM= Structural Equation Models. [a]Logit model were used. [b]Probit model were used, when appropriated.

	Direct			Indirect			Total		
	Bias	MSE	EmpSE	Bias	MSE	EmpSE	Bias	MSE	EmpSE
n = 1,000	Estimate±SE	Estimate±SE	Estimate±SE	Estimate±SE	Estimate±SE	Estimate±SE	Estimate±SE	Estimate±SE	Estimate±SE
4SLS[a]	-0.001±0.001	0.003±0.000	0.098±0.001	0.002±0.000	0.000±0.000	0.044±0.001	0.001±0.001	0.004±0.000	0.038±0.001
3SLS[a]	-0.001±0.001	0.003±0.000	0.098±0.001	-0.004±0.001	0.001±0.000	0.040±0.001	-0.004±0.001	0.004±0.000	0.034±0.001
4SRI[a]	-0.001±0.001	0.003±0.000	0.131±0.001	0.002±0.000	0.000±0.000	0.058±0.001	0.001±0.001	0.004±0.000	0.047±0.001
3SRI[a]	-0.001±0.001	0.003±0.000	0.131±0.001	-0.003±0.000	0.000±0.000	0.053±0.001	-0.004±0.001	0.004±0.000	0.043±0.001
SEM[b]	-0.012±0.002	0.004±0.000	0.060±0.007	-0.021±0.001	0.001±0.000	0.024±0.007	-0.001±0.000	0.000±0.000	0.027±0.001
n = 2,500									
4SLS[a]	-0.000±0.000	-0.001±0.000	0.062±0.000	0.004±0.000	0.000±0.000	0.026±0.001	0.004±0.000	0.002±0.000	0.022±0.001
3SLS[a]	-0.000±0.000	-0.001±0.000	0.062±0.000	-0.001±0.000	0.000±0.000	0.024±0.001	-0.002±0.000	0.002±0.000	0.020±0.001
4SRI[a]	-0.000±0.000	-0.001±0.000	0.082±0.000	0.004±0.000	0.000±0.000	0.033±0.000	0.004±0.000	0.002±0.000	0.028±0.001
3SRI[a]	-0.000±0.000	-0.001±0.000	0.082±0.000	-0.001±0.000	0.000±0.000	0.031±0.000	-0.002±0.000	0.002±0.000	0.026±0.001
SEM[b]	-0.012±0.001	0.002±0.000	0.038±0.000	-0.020±0.000	0.001±0.000	0.014±0.000	-0.001±0.000	0.000±0.000	0.014±0.001
n = 10,000									
4SLS[a]	-0.000±0.000	0.000±0.000	0.029±0.000	0.005±0.000	0.000±0.000	0.013±0.000	0.005±0.000	0.000±0.000	0.012±0.000
3SLS[a]	-0.000±0.000	0.000±0.000	0.029±0.000	-0.000±0.000	0.000±0.000	0.011±0.000	-0.000±0.000	0.000±0.000	0.011±0.000
4SRI[a]	-0.000±0.000	0.000±0.000	0.039±0.000	0.005±0.000	0.000±0.000	0.017±0.000	0.005±0.000	0.000±0.000	0.016±0.000
3SRI[a]	-0.000±0.000	0.000±0.000	0.039±0.000	-0.000±0.000	0.000±0.000	0.016±0.000	-0.000±0.000	0.000±0.000	0.015±0.000
SEM[b]	-0.011±0.000	0.000±0.000	0.019±0.000	-0.020±0.000	0.000±0.000	0.007±0.000	-0.001±0.000	0.000±0.000	0.006±0.000

Table 6. Bias, Mean Squared Error (MSE), and Empirical Standard Error (EmpSE) in Scenario S2 according to different sample sizes, based on coefficients shown in Table D1. MSE=Mean Squared Error; EmpSE=Empirical Standard Error; SE= standard error;4SLS=4-Stage-Least-Squares; 3SLS=3-Stage-Least-Squares; 4SRI=4-Stage-Residual-Inclusion; 3SRI=3-Stage-Residual-Inclusion; SEM= Structural Equation Models. [a]Logit model were used. [b]Probit model were used, when appropriated.

	Direct			Indirect			Total		
S3	**Bias**	**MSE**	**EmpSE**	**Bias**	**MSE**	**EmpSE**	**Bias**	**MSE**	**EmpSE**
n = 1,000	**Estimate±SE**	**Estimate±SE**	**Estimate±SE**	**Estimate±SE**	**Estimate±SE**	**Estimate±SE**	**Estimate±SE**	**Estimate±SE**	**Estimate±SE**
***4SLS*[b]**	-0.001±0.001	0.003±0.000	0.058±0.001	-0.014±0.002	0.007±0.000	0.086±0.001	-0.015±0.002	0.007±0.000	0.084±0.001
***3SLS*[b]**	-0.001±0.001	0.003±0.000	0.058±0.001	-0.018±0.002	0.006±0.000	0.083±0.001	-0.019±0.002	0.007±0.000	0.082±0.001
***4SRI*[b]**	-0.006±0.001	0.003±0.000	0.050±0.001	0.006±0.001	0.003±0.000	0.052±0.001	-0.000±0.002	0.006±0.000	0.076±0.001
***3SRI*[b]**	-0.006±0.001	0.003±0.000	0.050±0.001	0.000±0.001	0.003±0.000	0.051±0.001	-0.006±0.002	0.006±0.000	0.076±0.001
***SEM*[b]**	-0.021±0.010	0.193±0.106	0.439±0.007	-0.017±0.010	0.193±0.106	0.439±0.007	-0.038±0.001	0.004±0.000	0.054±0.001
n = 2,500									
***4SLS*[b]**	-0.000±0.001	0.001±0.000	0.003±0.000	-0.011±0.001	0.002±0.000	0.043±0.001	-0.011±0.001	0.002±0.000	0.047±0.001
***3SLS*[b]**	-0.000±0.001	0.001±0.000	0.003±0.000	-0.015±0.001	0.002±0.000	0.040±0.001	-0.015±0.001	0.002±0.000	0.046±0.001
***4SRI*[b]**	-0.004±0.001	0.001±0.000	0.030±0.000	0.007±0.001	0.001±0.000	0.033±0.000	0.003±0.001	0.002±0.000	0.047±0.001
***3SRI*[b]**	-0.004±0.001	0.001±0.000	0.030±0.000	0.002±0.001	0.001±0.000	0.032±0.000	-0.003±0.001	0.002±0.000	0.046±0.001
***SEM*[b]**	0.000±0.001	0.001±0.000	0.033±0.000	-0.035±0.000	0.001±0.000	0.014±0.000	-0.035±0.001	0.002±0.000	0.033±0.001
S4	**Bias**	**MSE**	**EmpSE**	**Bias**	**MSE**	**EmpSE**	**Bias**	**MSE**	**EmpSE**
n = 1,000	**Estimate±SE**	**Estimate±SE**	**Estimate±SE**	**Estimate±SE**	**Estimate±SE**	**Estimate±SE**	**Estimate±SE**	**Estimate±SE**	**Estimate±SE**
4SLS[a]	-0.007±0.005	0.024±0.001	0.155±0.001	-0.026±0.010	0.108±0.008	0.327±0.001	0.135±0.010	0.108±0.008	0.551±0.001
3SLS[a]	-0.007±0.005	0.024±0.001	0.155±0.001	-0.037±0.010	0.096±0.007	0.308±0.001	0.124±0.009	0.099±0.006	0.549±0.001
4SRI[a]	-0.003±0.005	0.023±0.001	0.153±0.001	-0.012±0.007	0.053±0.003	0.231±0.001	0.154±0.008	0.085±0.004	0.323±0.001
3SRI[a]	-0.003±0.005	0.023±0.001	0.152±0.001	-0.024±0.007	0.049±0.003	0.219±0.001	0.142±0.007	0.077±0.004	0.307±0.001
SEM[b]	-0.005±0.002	0.006±0.000	0.088±0.001	-0.042±0.001	0.003±0.000	0.036±0.007	-0.002±0.000	0.000±0.000	0.052±0.001
n = 2,500									
4SLS[a]	-0.001±0.003	0.009±0.000	0.095±0.000	-0.036±0.006	0.035±0.002	0.185±0.001	0.131±0.005	0.046±0.002	0.249±0.001
3SLS[a]	-0.001±0.003	0.009±0.000	0.095±0.000	-0.047±0.005	0.031±0.002	0.169±0.001	0.121±0.005	0.040±0.002	0.230±0.001
4SRI[a]	0.000±0.003	0.008±0.000	0.090±0.000	-0.007±0.004	0.015±0.001	0.122±0.000	0.163±0.004	0.045±0.002	0.132±0.001
3SRI[a]	0.000±0.003	0.008±0.000	0.090±0.000	-0.020±0.004	0.014±0.001	0.115±0.000	0.149±0.004	0.040±0.002	0.122±0.001
SEM[b]	-0.006±0.001	0.002±0.000	0.053±0.000	-0.042±0.000	0.002±0.000	0.018±0.000	-0.001±0.000	0.000±0.000	0.028±0.001

Table 7. Bias, Mean Squared Error (MSE), and Empirical Standard Error (EmpSE) in Scenarios S3 and S4 according to different sample sizes, based on coefficients shown in Table D1.
MSE=Mean Squared Error; EmpSE=Empirical Standard Error; SE= standard error;4SLS=4-Stage-Least-Squares; 3SLS=3-Stage-Least-Squares; 4SRI=4-Stage-Residual-Inclusion; 3SRI=3-Stage-Residual-Inclusion; SEM= Structural Equation Models. [a]Logit model were used. [b]Probit model were used, when appropriated.

	Direct			Indirect			Total		
S5	**Bias**	**MSE**	**EmpSE**	**Bias**	**MSE**	**EmpSE**	**Bias**	**MSE**	**EmpSE**
n = 1,000	**Estimate±SE**	**Estimate±SE**	**Estimate±SE**	**Estimate±SE**	**Estimate±SE**	**Estimate±SE**	**Estimate±SE**	**Estimate±SE**	**Estimate±SE**
4SLS[b]	-0.027±0.068	4.647±0.329	2.156±0.001	0.113±0.1000	10.005±0.827	3.162±0.001	0.904±0.108	12.532±0.825	10.720±0.001
3SLS[b]	-0.027±0.068	4.647±0.329	2.156±0.001	-0.027±0.097	9.493±0.784	3.082 ±0.001	0.764±0.106	11.889±0.783	10.185±0.001
4SRI[b]	0.026±0.082	6.750±0.461	2.599±0.001	0.204±0.152	23.180±2.237	4.813±0.001	1.049±0.157	25.876±2.058	11.450±0.001
3SRI[b]	0.026±0.082	6.750±0.461	2.599±0.001	0.025±0.149	22.194±2.114	4.712±0.001	0.87±0.155	24.685±1.923	12.575±0.001
SEM[b]	-0.110±0.014	0.211±0.010	0.447±0.002	-0.128±0.005	0.040±.003	0.255±0.002	-0.033±0.003	0.011±0.002	0.119±0.001
n = 2,500									
4SLS[b]	-0.085±0.037	1.376±0.066	1.170±0.000	-0.189±0.044	1.996±0.145	1.401±0.001	0.544±0.054	3.184±0.185	2.130±0.001
3SLS[b]	-0.085±0.037	1.376±0.066	1.170±0.000	-0.314±0.042	1.840±0.132	1.320±0.001	0.419±0.052	2.893±0.165	2.212±0.001
4SRI[b]	-0.080±0.046	2.124±0.116	1.456±0.000	-0.208±0.063	3.993±0.293	1.988 ±0.000	0.531±0.074	5.739±0.327	3.347±0.001
3SRI[b]	-0.080±0.046	2.124±0.116	1.456±0.000	-0.389±0.060	3.697±0.268	1.884±0.000	0.350±0.071	5.222±0.294	3.284±0.001
SEM[b]	-0.110±0.008	0.077±0.004	0.154±0.001	-0.126±0.002	0.019±0.000	0.061±0.001	-0.020±0.001	0.001±0.000	0.018±0.001
S6	**Bias**	**MSE**	**EmpSE**	**Bias**	**MSE**	**EmpSE**	**Bias**	**MSE**	**EmpSE**
n = 1,000	**Estimate±SE**	**Estimate±SE**	**Estimate±SE**	**Estimate±SE**	**Estimate±SE**	**Estimate±SE**	**Estimate±SE**	**Estimate±SE**	**Estimate±SE**
4SLS[a]	-0.107±0.032	2.026±0.331	1.420±0.022	0.023±0.033	2.161±0.373	1.470±0.023	-0.084±0.039	2.998±0.353	1.730±0.027
3SLS[a]	-0.107±0.032	2.026±0.331	1.420±0.022	-0.014±0.032	2.098±0.362	1.448±0.023	-0.121±0.039	3.052±0.365	1.743±0.027
4SRI[a]	-0.268±0.032	2.972±0.124	1.415±0.022	-0.119±0.036	2.666±0.282	1.629±0.026	-0.388±0.039	3.292±0.189	1.773±0.028
3SRI[a]	-0.268±0.032	2.972±0.124	1.415±0.022	-0.130±0.036	2.598±0.027	1.607±0.025	-0.398±0.039	3.260±0.184	1.764±0.028
SEM[b]	0.004±0.005	0.049±0.002	0.223±0.003	-0.119±0.002	0.026±0.001	0.109±0.002	-0.116±0.006	0.081±0.003	0.260±0.004
n = 2,500									
4SLS[a]	-0.046±0.013	0.366±0.020	0.603±0.009	-0.004±0.011	0.225±0.015	0.475±0.007	-0.050±0.016	0.533±0.039	0.729±0.011
3SLS[a]	-0.046±0.013	0.366±0.020	0.603±0.009	-0.038±0.010	0.211±0.015	0.458±0.007	-0.084±0.016	0.536±0.044	0.727±0.011
4SRI[a]	-0.251±0.016	0.601±0.027	0.738±0.012	-0.108±0.012	0.335±0.021	0.569±0.009	-0.358±0.019	0.875±0.043	0.866±0.013
3SRI[a]	-0.251±0.016	0.601±0.027	0.738±0.012	-0.118±0.012	0.318±0.021	0.551±0.009	-0.369±0.019	0.875±0.047	0.860±0.014
SEM[b]	-0.009±0.003	0.017±0.001	0.129±0.002	-0.120±0.001	0.017±0.000	0.054±0.001	-0.130±0.003	0.037±0.001	0.144±0.002

Table 8. Bias, Mean Squared Error (MSE), and Empirical Standard Error (EmpSE) in Scenarios S5 and S6 according to different sample sizes, based on coefficients shown in Table D1.
MSE=Mean Squared Error; EmpSE=Empirical Standard Error; SE= standard error;4SLS=4-Stage-Least-Squares; 3SLS=3-Stage-Least-Squares; 4SRI=4-Stage-Residual-Inclusion; 3SRI=3-Stage-Residual-Inclusion; SEM= Structural Equation Models. [a]Logit model were used. [b]Probit model were used, when appropriated.

5.5.2. Application to real data: The PanGenEU case-control study

Considering the performance of the five approaches, we applied the appropriate methodology to the PanGenEU study.

We defined the Biological Model A, where the aim was to study the causal effect of BMI (continuous) on PC risk (binary, case/control) considering LSDM (binary) as mediator, which corresponds to scenario S4, the Biological Model B, to study the causal effect of obesity (binary) on PC risk also considering LSDM as mediator, scenario S6, and the Biological Model C, to analyse the causality of LSDM on PC considering obesity as mediator, scenario S6. We included those participants with available genetic information, (1,040 cases and 738 controls). More details are provided in Appendix B, in Molina-Montes et al. (2021) and López de Maturana et al. (2021).

The method selected to estimate each effect was based on the results shown in Tables 7 and 8. Based on the good performance values of bias, MSE, and coverage values, we applied the 3SRI for Biological Model A and 4SLS for Biological Models B and C. Results are shown in Table 9, also reporting those previously obtained in Molina-Montes et al. (2021), to compare *MRinCMA* with a classic CMA. We conclude that, unlike CMA, after adjusting by genetic scores, we did not find any evidence of causality.

Biological model A			**Direct Effect**		**Indirect Effect**		**Total effect**	
Exposure	**Mediator**	**Method**	**OR**	**CI95%(OR)**	**OR**	**CI95%(OR)**	**OR**	**CI95%(OR)**
BMI at age 50	*LSDM*	3SRI	1.02	[0.84;1.22]	1.21	[0.81;1.95]	1.00	[0.94;1.05]
Biological model B			**Direct Effect**		**Indirect Effect**		**Total effect**	
Exposure	**Mediator**	**Method**	**OR**	**CI95%(OR)**	**OR**	**CI95%(OR)**	**OR**	**CI95%(OR)**
Obesity at age 50	*LSDM*	4SLS	0.97	[0.62;1.53]	1.13	[0.89;1.88]	1.12	[0.72;1.65]
		CMA[a]	0.88	[0.72;1.07]	1.07	[1.04;1.15]	0.94	[0.78;1.14]
Biological model C			**Direct Effect**		**Indirect Effect**		**Total effect**	
Exposure	**Mediator**	**Method**	**OR**	**CI95%(OR)**	**OR**	**CI95%(OR)**	**OR**	**CI95%(OR)**
LSDM	*Obesity at age 50*	4SLS	1.20	[0.87;1.64]	1.00	[0.87; 1.12]	1.19	[0.83;1.73]
		CMA[a]	1.53	[1.13;2.09]	0.99	[0.87;1.07]	1.5	[1.14;2.17]

Table 9. Direct, indirect and total effects and 95%CI from the PanGenEU case-control study
BMI: Body Mass Index; LSDM: long-standing diabetes mellitus; 3SRI: 3-Stage-Residual-Inclusion; 4SLS: 4-Stage-Least-Squares; CI: Confidence Interval. Confidence Interval were calculated using bootstrap, considering 2.5% and 97.5% percentiles. Logit models were considered. [a]CMA: Causal Mediation Analysis previously obtained in Molina-Montes et al. (2021).

5.6. Discussion

In this chapter we propose an extension of CMA and MR approaches (*MRinCMA)* to relax the potential confounding bias assumptions in the CMA by incorporating two *IV*s, which can deal with continuous and categorical variables. It is important to highlight that the 4SLS, 4SRI and 3SLS approaches, were already proposed but only for continuous and normally distributed variables. Moreover, we propose a 3SRI approach using three stages considering observed and residual values rather than predicted values. We showed that *MRinCMA* provided unbiased estimates of the direct, indirect, and total effects among most of the simulations considered.

In the first scenario (S1) where E, M, and Y were defined as continuous variables, we conclude that *MRinCMA* worked properly, obtaining unbiased results equivalent or better than SEM, in all simulation settings. Our results, in terms of bias, are similar to those obtained by Burgess et al. (2015) when they considered 3SLS and SEM. Frölich et al. (2017) also confirmed that 4SLS worked properly in several simulation scenarios.

In the other scenarios, where we have at least one categorical variable, as it was already described by Burgess (2013) and by Carter et al. (2021), these approaches can lead to less precise estimations due to the non-collapsibility of the odds-ratio, resulting in higher bias. Even so, in these cases, *MRinCMA* always outperformed SEM, where worst coverage rates were obtained. Up to now, only Carter et al. (2021) and our proposal provide unbiased estimates with non-continuous variables, but our proposal used the residual values instead of predicted values, an approach that has been described to work better for categorical variables (Burgess, Daniel, et al. 2015; Frölich and Huber 2017; Terza et al. 2008).

Some issues regarding the estimation of the standard errors (SE) should be mentioned. As it was already described (Palmer et al. 2011, 2017; Terza et al. 2008), approaches that involve several stages underestimates the SE and consequently this effect would impact on the performance of *MRinCMA*. To avoid this limitation, the SE and the 95%CI were derived using bootstrapping.

This work represents an exhaustive analysis, where different *MRinCMA* approaches were tested in several settings. We proved that *MRinCMA* can also be used when the variables of interest are non-continuous, as is often the case in epidemiological studies. Considering the range of simulations implemented, we also believe this project could be considered as a guideline for investigators that are interested in applying CMA incorporating instrumental variables. However, some possible extensions of these methods could be explored in further studies. For example, we

did not consider a horizontal pleiotropy affecting the *IV* (Bowden et al. 2015; Burgess and Thompson 2017; Carter et al. 2021; Rees et al. 2017; Sanderson 2021). Moreover, it could be interesting to study the possible weaknesses of the various methods when the IV assumptions do not hold (Burgess, Daniel, et al. 2015; Frölich and Huber 2017; North et al. 2019). Lastly, in our study we did not consider a possible interaction between E and M (Burgess, Daniel et al. 2015; Frölich and Huber 2017; North et al. 2019).

5.7. Conclusion

This study provides new methods to address CMA interrogations by incorporating genetic information related to exposure and mediator to correct confounding bias. *MRinCMA* can be easily applied in a wide range of epidemiological and clinical scenarios, regardless the nature of the variables and it could be considered as a solution in those studies where the main objective is to apply CMA, but confounding assumptions do not hold.

Chapter 6: Stratification in Mendelian Randomization

Chapter 6

In this chapter I present a new approach proposed to avoid potential collider bias when Mendelian Randomization analyses are stratified by a collider variable. This new extension allows researchers to conduct MR in subgroups of the population while avoiding introducing bias.

6.1. Abstract

Background

Mendelian randomization (MR) uses genetic variants as instrumental variables to investigate the causal effect of a risk factor on an outcome. A collider is a variable influenced by two or more other variables. Naive calculation of MR estimates in strata of the population defined by a collider, such as a variable affected by the risk factor can result in collider bias.

Methods

We propose an approach that allows MR estimation in strata of the population while avoiding collider bias. This approach constructs a new variable, the residual collider, as the residual from regression of the collider on the genetic instrument, and then calculates causal estimates in strata defined by quantiles of the residual collider. Estimates stratified on the residual collider will typically have an equivalent

interpretation to estimates stratified on the collider, but they are not subject to collider bias. We apply the approach in several simulation scenarios considering different characteristics of the collider variable and strengths of the instrument. We then apply the proposed approach to investigate the causal effect of smoking on bladder cancer in strata of the population defined by bodyweight.

Results

The new approach generated unbiased estimates in all the simulation settings. In the applied example, we observed a trend in the stratum-specific MR estimates at different bodyweight levels that suggested stronger effects of smoking on bladder cancer among individuals with lower bodyweight.

Conclusions

The proposed approach can be used to perform MR studying heterogeneity among subgroups of the population while avoiding collider bias.

Publication

We present the paper that has been submitted to the European Journal of Epidemiology, which is now under the second revision.
Published in MedRxiv, August 2021:
https://www.medrxiv.org/content/10.1101/2021.08.17.21262178v1

6.2. Introduction

Mendelian randomization (MR) is the use of genetic variants as instrumental variables to assess the causal relationship between a risk factor and an outcome (Burgess et al. 2015; Haycock et al. 2016). A valid instrumental variable (IV), or genetic instrument, must meet the following assumptions (Burgess et al. 2016): IV1, the instrument is associated with the risk factor; IV2, the instrument cannot affect the outcome directly, only potentially indirectly via the risk factor; and IV3, the instrument is not associated with any measured or unmeasured confounders (Figure 14A). If these assumptions are satisfied, an association of the instrument with the outcome is indicative of a causal effect of the risk factor on the outcome (Haycock et al. 2016; Hernán and Robins 2006). For point estimation of a causal effect, a further parametric assumption (known as IV4) is required. Two common assumptions are 1) monotonicity: the effect of the IV on the exposure is in the same direction (either an increase or a decrease) for all individuals in the population; or 2) homogeneity: a sufficient assumption is that the causal effect of the exposure on the outcome is constant for all individuals in the population (Hernán and Robins 2006). Under monotonicity, the IV estimate can be interpreted as a local average causal effect; under homogeneity, it can be interpreted as an average causal effect (Swanson and Hernán 2013). If either the IV2 or IV3 assumption is not satisfied, then the instrument could be associated with the outcome in the absence of a causal effect of the risk factor. However, only the IV1 assumption can be verified based on measured data (Bowden et al. 2016).

Collider bias can occur when conditioning on a collider, defined as a variable that is a common effect of two or more variables (Hernán et al. 2004; Munafò et al. 2018; Paternoster et al. 2017; Pearl 1995). The existence of a collider can be recognized in a causal diagram when there are two arrows pointing at the same variable; the node at which the arrowheads "collide" together is a collider. For example, in the standard MR diagram, the risk factor is a collider as it is affected by both the instrument and the confounders. Moreover, any variable that is a causal descendent of a collider is also affected by the same variables and so is itself a collider; hence in MR any variable influenced by the risk factor is a collider (Figure 14B). Even if the variables influencing a collider are independent, they will typically become dependent when conditioning on the collider. Hence conditioning on a variable affected by the risk factor will typically generate a conditional association between the instrument and the confounders, violating the IV3 assumption, and biasing Mendelian randomization estimates of the risk factor on the outcome.

Selection bias is a form of collider bias that occurs when selection of individuals into a dataset is dependent on a collider. For example, when disease progression is considered as an outcome, only patients who have already developed the disease would be recruited into the study (Paternoster et al. 2017). If the risk of developing the disease is influenced by the risk factor, then it is a collider when considering disease progression as the outcome, and selection of the study sample would result in collider bias. Several papers related to selection bias in the context of IV analysis and MR have been already published (Boef et al. 2015; Canan et al. 2017; Gkatzionis and Burgess 2019; Hughes et al. 2019; Smit et al. 2019). Inverse probability weighting (IPW) on the probability of selection has been proposed as a method to avoid selection bias (Canan et al. 2017; Gkatzionis and Burgess 2019).

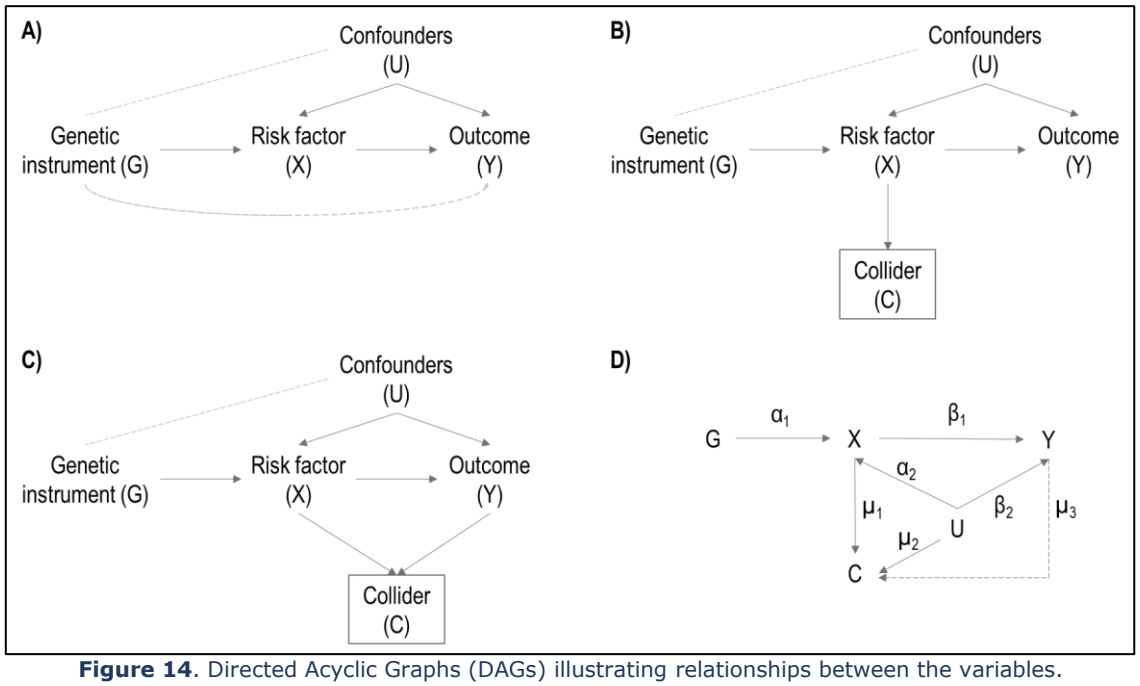


Figure 14. Directed Acyclic Graphs (DAGs) illustrating relationships between the variables.

A) Mendelian Randomization causal diagram with the instrumental variable assumptions. The dashed lines between G and Y and between G and U, represent violations of the *IV*2 and *IV*3 assumptions respectively.
B) DAG considering a collider variable C, being a common child of genetic instrument G and confounders U. When conditioning on C (indicated by the square box on C), G and U become correlated (dashed line between G and U) and a violation of the *IV*3 assumption occurs.
C) DAG considering a collider variable C, being a common child of risk factor X and outcome Y.
D) DAG illustrating the variables and parameters used for the simulation study. Dash line from Y to C correspond to simulation scenarios B1 to B3

Collider bias could also occur when stratifying the population based on a collider. As an example, we consider investigating the causal effect of the risk factor on the outcome for individuals with specific levels of a stratifying variable. Stratification is important for identifying whether there are subgroups of the population for which causal effects of the risk factor are different, and so the outcome would be affected

more strongly by an intervention on the risk factor. However, if the stratifying variable is a collider, an association between the instrument and the outcome in strata of the population could arise due to collider bias, invalidating the results. In particular, collider bias could affect some estimates more than others, leading to heterogeneity in the stratum-specific causal estimates even if the true causal effect is the same across strata. Although several previous papers have considered collider bias arising due to differential selection into the study sample (Canan et al. 2017; Gkatzionis and Burgess 2019; Swanson 2019), including when selection is driven by differential survival (a specific example of collider bias known as survival bias) (Hu and Mustillo 2016; Martinussen et al. 2017; Tchetgen et al. 2015), we are not aware of previous work considering the impact of stratification on a collider variable. IPW is unlikely to be a viable solution in this case, as the stratifying variable is typically continuous.

The aim of this paper is to present an MR approach that obtains estimates in strata of the population that do not suffer from collider bias. The structure of this paper is as follows: first, we demonstrate the bias that arises from conditioning on a collider; second, we propose an approach to calculate MR estimates in strata of the population and evaluate heterogeneity between estimates in the different strata; third, we illustrate this new technique in simulation studies and an applied example using the UK Biobank resource; and finally, we discuss the interpretation of estimates and limitations of the approach.

6.3. Methods

6.3.1. Illustration of collider bias

The simplest MR method to estimate the causal effect of a risk factor X on outcome Y with a genetic instrument G is the ratio method (S Burgess et al. 2017). With a single instrument, a continuous risk factor and outcome, and under assumptions of linearity and no effect modification, the ratio estimate previously defined in [23] is: $\hat{\theta} = \frac{\hat{\beta}_{YG}}{\hat{\beta}_{XG}}$, where $\hat{\beta}_{YG}$ the coefficient from regressing Y on G, and $\hat{\beta}_{XG}$ is the coefficient from regressing X on G (Burgess, Small et al. 2017). If data on G, X, and Y are available in the same individuals (known as "one-sample MR"), the same estimate with a single IV can be obtained using the two-stage-least-squares method.

Collider bias will occur when adjusting for a collider variable C in the regression models for the ratio estimate, since an association between the instrument and the

outcome will occur through conditioning on the collider. To demonstrate the impact and magnitude of collider bias, we performed a simulation study in which we compared estimates when no adjustment on C is made versus when the outcome regression is adjusted for C. It is also possible to adjust the risk factor regression for C; however, while this will distort estimates, this adjustment alone will not bias causal estimates when the true causal effect is null.

6.3.2. Stratification in Mendelian randomization

To further illustrate the impact of collider bias, we performed a simulation study in which we calculated causal estimates using the ratio method within strata of the population defined using a variable that is influenced by the risk factor, and hence is a collider. We compared two approaches: first, we stratified directly on the collider C, and second, we stratified on a new variable C_0, referred to as the "residual collider". The residual collider was generated as the residual from regression of the collider on the genetic instruments:

$C_0 = C - \hat{C}$, where $\hat{C}$ are the fitted values from regression of C on G.

The residual collider C_0 is not associated with the instrument, and hence it is not itself a collider. It is influenced by the component of the risk factor that is not a function of G, but not by the component that is a function of G. However, provided that the genetic instrument does not explain much of the variance in the risk factor (as is typical in a MR application), it is likely not to explain much of the variance in the collider, and so the residual collider will be highly correlated with the collider. Hence, while stratifying on the residual collider is important to avoid bias, the strata defined by stratifying on the collider or residual collider are likely to be similar and so any difference in the interpretation of stratum-specific estimates is minimal. If the genetic instrument explains a substantial portion of variance in the risk factor, then the residual collider will not be as highly correlated with the collider.

Here we considered estimates in four strata of the population defined by quartiles of the distribution of the collider or residual collider; however, in practice any number of strata could be considered. We estimated genetic associations with the outcome in each stratum separately. We estimated genetic associations with the risk factor in the full dataset, although if it is believed that these associations vary between strata, it would be possible to estimate these within each stratum as well. The stratum-specific estimate is calculated as the ratio of the stratum-specific genetic association with the outcome divided by the genetic association with the risk factor. The interpretation of stratum-specific estimates is equivalent to that of IV estimates

obtained in the whole population; depending on the version of the IV4 assumption, they either target an average or a local average causal effect. We also investigated heterogeneity between the stratum-specific estimates using Cochran's Q statistic (Higgins et al. 2003), and (in the applied example) we examined the presence of a trend in the estimates by meta-regression of the stratum-specific estimates on the median value of the collider in each stratum (Thompson and Higgins 2002).

6.4. Simulation set-up

To investigate the impact of collider bias in realistic scenarios, we generated simulated data using the following data-generating model:

$$G, U, \varepsilon_X, \varepsilon_Y, \varepsilon_C \sim N(0,1) \text{ independently}$$

$$X = \alpha_0 + \alpha_1 G + \alpha_2 U + \varepsilon_X \quad [38]$$

$$Y = \beta_0 + \beta_1 X + \beta_2 U + \varepsilon_Y \quad [39]$$

$$C = \mu_0 + \mu_1 X + \mu_2 U + \varepsilon_C \quad [40]$$

We simulated the instrument G, the confounder U, and the error terms for X, Y and C, ε_X, ε_Y *and* ε_C, as independent normally distributed variables. The risk factor X is defined as a linear combination of the instrument, the confounder, and the error term. The outcome Y and the collider C are both linear combinations of the risk factor, confounder, and their error terms. In each simulated dataset, we also generated the residual collider C_0 as the residual from regression of C on G as previously described.

The causal estimate of interest is β_1, while α_2 and β_2 represent the effects of U on X and Y respectively; α_1 is the effect of G on X; and μ_1 and μ_2 are the effects of X and U on C, respectively.

We considered three scenarios based on the parameter β_1: Scenario A1, where there is a null causal effect of X on Y $(\beta_1 = 0)$; Scenario A2, where the effect is constant and positive $(\beta_1 = 0.5)$; and Scenario A3, where the effect depends on the collider $(\beta_1 = 0.5 + 0.2C)$. In Scenario A1, we considered estimates from the ratio method with and without adjustment for the collider. In Scenarios A2 and A3, we consider stratum-specific estimates from stratification on the collider C or the residual collider C_0.

We varied the other parameters to consider the impact of different settings on collider bias:

i) $\alpha_1 = (0.05, 0.1, 0.3)$, in order to study the impact of the strength of the instrument on estimates.

ii) Positive confounding $(\alpha_2 = 0.8,\ \beta_2 = 0.8)$, negative $(\alpha_2 = -0.8,\ \beta_2 = -0.8)$ and mixed $(\alpha_2 = 0.8,\ \beta_2 = -0.8)$, to study how the direction of confounding affects the estimates.

iii) μ_1 and $\mu_2 = (-1, -0.5, 0, 0.5, 1)$ to study how the strength of the collider effects influences bias.

We also considered scenarios where the collider is a common effect of X and Y (Figure 14C). In these scenarios, the collider is generated as:

$C = \mu_0 + \mu_1 X + \mu_2 U + \mu_3 Y + \varepsilon_C$, where $\mu_2 = 0.3$ and $\mu_3 = (-1, -0.5, 0, 0.5, 1)$.

In Scenario B1, the causal effect of X on Y is null $(\beta_1 = 0)$, in Scenario B2, the causal effect is constant and positive $(\beta_1 = 0.5)$, and in Scenario B3, the causal effect depends on U $(\beta_1 = 0.5 + 0.2U)$, as it is not possible for the causal effect to depend on C when C is a function of Y.

Finally, we investigated additional scenarios with a binary outcome Y. We generate Y from a Binomial distribution where the probability is obtained from a logit transformation as:

$\text{logit}\left(P(Y=1)\right) = \beta_0 + \beta_1 X + \beta_2 U$, where $\beta_0 = 0.5$.

In Scenario C1, the causal effect of X on Y is null $(\beta_1 = 0)$, in Scenario C2, the causal effect is constant and positive $(\beta_1 = 0.5)$ and in Scenario C3, the causal effect depends on C $(\beta_1 = 0.5 + 0.2C)$.

In the binary outcome scenarios, genetic associations with the outcome were estimated by logistic regression. For these additional scenarios, we only consider

$\alpha_1 = 0.1$ and the positive confounding values; otherwise, we consider all parameters as in scenarios A1 to A3.

We considered a sample size of $n = 10{,}000$ and $m = 500$ replications for each set of parameter values. A directed acyclic graph illustrating the simulation parameters is shown in Figure 13D.

6.5. Applied example: effect of tobacco smoking on bladder cancer risk across bodyweight strata

We applied the proposed MR stratification approach to investigate the causal effect of tobacco smoking on bladder cancer across strata of the population defined by bodyweight. Tobacco smoking is one of the strongest risk factors for cancer, and it has already been reported to be causally associated with bladder cancer risk in a previous Mendelian randomization study (Larsson et al. 2020). With our current example, the objective was to investigate whether the effect of smoking on the risk of developing bladder cancer is homogeneous across the bodyweight distribution of the population, while avoiding potential collider bias by applying our new stratification approach.

We performed analyses in the UK Biobank study, a population-based cohort of more than 500,000 United Kingdom residents recruited between 2006 and 2010 (Sudlow et al. 2015). For our analysis, we restricted to unrelated European ancestry participants, resulting in a final sample size of 367,643 individuals following sample selection and quality control procedures as described previously (Larsson et al. 2020). The risk factor is a binary variable representing the smoking behaviour, defined as being a current smoker versus a former or never smoker; the stratifying variable is bodyweight, measured in kg; and the binary outcome is bladder cancer status, defined based on the data from national registries (International Classification of Diseases 9th edition codes: 188, 189.1, 189.2, V10.51, V10.53; or International Classification of Diseases 10th edition codes: C67, C65, C66, Z85.51, Z85.54, Z85.53), and self-reported information from an interview with a nurse practitioner. The instrument for smoking was a weighted genetic risk score comprising 378 conditionally independent SNPs obtained from a genome-wide association study (GWAS) assessing associations with smoking initiation (i.e., probability of ever smoked regularly), and weighted by the associations with smoking initiation (Liu et al. 2019). Genetic associations with the risk factor and outcome were obtained by logistic regression in UK Biobank with adjustment for age, sex, and 10 genomic principal components. While age, sex, and principal components cannot logically be

colliders as they are not affected by the risk factor or outcome, bodyweight is likely to be a collider, as it is influenced by smoking status (Taylor et al. 2019).

6.6. Results

6.6.1. Illustration of collider bias

Results from Scenario A1 ($\beta_1 = 0$, null causal effect) are presented in Table 10 for $\alpha_1 = 0.1$ (corresponding to R^2=0.006 for the mean proportion of variance in the risk factor explained by the instrument and a mean F statistic of 60.8) and Tables F1 and F2 (Appendix F), for $\alpha_1 = 0.3$ (corresponding to R^2=0.051, mean F statistic of 548.6) and $\alpha_1 = 0.05$ (corresponding to R^2=0.001, mean F statistic of 15.3). In each case, we report the median estimate of β_1 across simulations, and the empirical type I error rate, representing the proportion of simulated datasets where the 95% confidence interval for the ratio estimate excludes zero. With no adjustment for the collider, median estimates were close to zero and empirical type I error rate was close to the expected value of 5%. When adjusting for the collider in the regression of Y on G, estimates were biased, and type I error rates were substantially above 5%. The only exception was for $\mu_1 = 0$; in this case, the variable C is not a function of the risk factor, and so does not act as a collider. Bias and type I error rates generally increased for more extreme values of μ_1 and μ_2 (both positive and negative values). The direction of bias depended on μ_1 and μ_2 and the direction of confounding.

6.6.2. Stratification in Mendelian randomization

Results from Scenario A2 ($\beta_1 = 0.5$, constant positive effect) are presented in Table 11 for $\alpha_1 = 0.1$ with positive confounding. Table F3 shows results for $\alpha_1 = 0.1$ with negative and mixed confounding, and Table F4 and F5, for $\alpha_1 = 0.3$ and $\alpha_1 = 0.05$. We report the median estimate of β_1 in four strata of the sample defined by quartiles of the collider C or residual collider C_0, and the proportion of simulated datasets for which the heterogeneity test statistic is rejected. When stratifying on the collider, median estimates were somewhat variable between the strata, although the proportion of datasets in which the heterogeneity test rejects the null hypothesis of homogeneity was not much above 5% in any scenario, reaching a maximum of 11% when $\alpha_1 = 0.3$. However, if we considered stronger instruments or larger sample

sizes, we would see this proportion considerably exceed 5% (see Table F6 where we first set $\alpha_1 = 0.5$ and n=10,000, and then set $\alpha_1 = 0.1$ and n = 50,000, and the type I error rate reached 16% in each case). This was due to increased precision of estimates; the magnitude of bias did not depend strongly on instrument strength. Median estimates differed substantially from the true value of 0.5 across strata, especially when the collider was strongly affected by the risk factor. In contrast, when stratifying on the residual collider, median estimates of β_1 were close to 0.5 throughout, and there was no suggestion in any case that the heterogeneity test rejected the null above the expected 5% rate.

Results from Scenario A3 (variable effect) are presented in Table 12 for $\alpha_1 = 0.1$ with positive confounding. Table F7 shows results for $\alpha_1 = 0.1$ with negative and mixed confounding, and Table F8 and F9 for $\alpha_1 = 0.3$ and $\alpha_1 = 0.05$. Estimates differed somewhat when stratifying on the collider versus the residual collider, although in both cases median estimates increased across the four strata. The proportion of datasets in which the heterogeneity test was rejected, which in this case represents the empirical power to detect heterogeneity in the stratum-specific estimates, was consistently higher when stratifying on the residual collider, indicating that true differences in the stratum-specific estimates were better detected when stratifying on the residual collider.

		Positive confounding (α_2 and β_2 = 0.8)				Negative confounding (α_2 and β_2 = -0.8)				Mixed confounding (α_2 =0.8 and β_2 = -0.8)			
		Median estimate	Type I error rate (%)	Median estimate	Type I error rate (%)	Median estimate	Type I error rate (%)	Median estimate	Type I error rate (%)	Median estimate	Type I error rate (%)	Median estimate	Type I error rate (%)
μ_1	μ_2	*No adjust for collider*		*Adjust Y/G for collider*		*No adjust for collider*		*Adjust Y/G for collider*		*No adjust for collider*		*Adjust Y/G for collider*	
-1	-1	0.01	7%	-0.27	70%	0.01	5%	0.09	10%	0.00	5%	0.28	69%
	-0.5	0.00	6%	-0.28	66%	-0.01	3%	-0.12	15%	0.00	6%	0.29	69%
	0	0.00	5%	-0.24	50%	-0.01	4%	-0.26	53%	0.01	5%	0.25	54%
	0.5	0.01	6%	-0.10	15%	0.01	6%	-0.28	69%	0.00	6%	0.12	15%
	1	0.00	5%	0.08	8%	0.01	3%	-0.27	68%	0.00	4%	-0.08	11%
-0.5	-1	0.01	6%	-0.16	30%	0.00	6%	0.15	21%	0.00	5%	0.17	34%
	-0.5	0.00	6%	-0.18	32%	0.00	5%	0.04	7%	0.01	4%	0.18	30%
	0	0.00	5%	-0.11	16%	0.00	3%	-0.12	14%	0.00	3%	0.11	14%
	0.5	-0.01	5%	0.02	5%	0.00	6%	-0.17	30%	0.00	4%	-0.03	7%
	1	0.01	5%	0.15	25%	0.00	6%	-0.18	34%	0.01	4%	-0.15	21%
0	-1	-0.01	7%	0.00	6%	0.00	6%	0.00	6%	0.00	6%	0.00	6%
	-0.5	0.00	7%	0.00	6%	0.00	6%	0.00	6%	0.00	5%	-0.01	5%
	0	0.00	6%	0.01	6%	0.01	6%	0.01	6%	0.00	6%	0.00	6%
	0.5	-0.01	6%	0.00	6%	0.00	5%	0.00	6%	0.00	7%	0.00	7%
	1	0.00	5%	0.01	6%	0.00	4%	0.00	4%	-0.01	6%	-0.01	5%
0.5	-1	0.01	4%	0.16	22%	0.01	6%	-0.17	35%	0.00	4%	-0.15	23%
	-0.5	-0.01	5%	0.02	5%	-0.01	7%	-0.19	36%	0.01	6%	-0.03	7%
	0	0.00	5%	-0.11	14%	0.00	4%	-0.11	15%	-0.01	4%	0.10	15%
	0.5	0.01	4%	-0.17	28%	0.00	5%	0.03	7%	0.00	5%	0.18	34%
	1	0.01	6%	-0.17	31%	0.01	5%	0.15	23%	0.00	5%	0.17	33%
1	-1	0.01	5%	0.08	8%	0.01	5%	-0.27	70%	0.01	3%	-0.07	9%
	-0.5	0.01	4%	-0.10	13%	0.01	5%	-0.27	64%	-0.01	5%	0.11	18%
	0	0.01	6%	-0.24	52%	0.00	5%	-0.24	50%	-0.01	3%	0.24	48%
	0.5	0.01	5%	-0.27	66%	0.01	4%	-0.11	15%	0.00	4%	0.28	66%
	1	0.01	4%	-0.26	68%	0.00	4%	0.08	10%	0.02	5%	0.29	75%

Table 10. Median of β_1 estimates and empirical Type I error rates for Scenario A1 (null causal effect, β_1 =0) with positive, negative, and mixed confounding, and α_1 =0.1. Empirical Type I error rate represents the proportion of simulated datasets where the null hypothesis is not rejected

		Positive confounding (α_2 and β_2 = 0.8)									
		Stratifying on collider, C					Stratifying on residual collider, C_0				
μ_1	μ_2	Proportion homogeneity rejected (%)	Median estimates Q1	Median estimates Q2	Median estimates Q3	Median estimates Q4	Proportion homogeneity rejected (%)	Median estimates Q1	Median estimates Q2	Median estimates Q3	Median estimates Q4
-1	-1	8%	0.11	-0.01	0.01	0.10	8%	0.49	0.48	0.49	0.50
	-0.5	6%	0.09	-0.05	-0.05	0.09	5%	0.53	0.48	0.51	0.53
	0	7%	0.07	0.01	-0.04	0.07	6%	0.50	0.52	0.50	0.49
	0.5	7%	0.19	0.10	0.09	0.17	6%	0.49	0.50	0.49	0.48
	1	4%	0.41	0.37	0.40	0.40	4%	0.50	0.48	0.52	0.49
-0.5	-1	7%	0.30	0.21	0.23	0.26	5%	0.53	0.53	0.52	0.48
	-0.5	5%	0.24	0.16	0.18	0.25	5%	0.48	0.47	0.51	0.48
	0	4%	0.29	0.23	0.21	0.29	4%	0.50	0.47	0.45	0.49
	0.5	6%	0.47	0.42	0.47	0.46	6%	0.50	0.50	0.50	0.50
	1	3%	0.59	0.65	0.64	0.60	5%	0.48	0.50	0.51	0.50
0	-1	6%	0.50	0.50	0.48	0.48	6%	0.51	0.50	0.48	0.47
	-0.5	4%	0.48	0.48	0.52	0.50	4%	0.47	0.47	0.52	0.50
	0	4%	0.49	0.48	0.51	0.52	4%	0.49	0.49	0.51	0.51
	0.5	5%	0.52	0.49	0.49	0.53	6%	0.54	0.49	0.48	0.53
	1	4%	0.48	0.48	0.51	0.49	4%	0.48	0.47	0.52	0.50
0.5	-1	4%	0.62	0.63	0.66	0.63	4%	0.51	0.49	0.52	0.52
	-0.5	5%	0.45	0.44	0.44	0.47	4%	0.49	0.50	0.49	0.51
	0	4%	0.32	0.23	0.23	0.29	4%	0.51	0.51	0.47	0.49
	0.5	5%	0.25	0.18	0.19	0.25	3%	0.51	0.49	0.52	0.50
	1	5%	0.26	0.23	0.20	0.28	5%	0.46	0.51	0.49	0.50
1	-1	5%	0.41	0.37	0.34	0.40	5%	0.49	0.50	0.46	0.49
	-0.5	5%	0.16	0.11	0.09	0.21	4%	0.49	0.49	0.50	0.51
	0	6%	0.12	-0.03	0.00	0.11	6%	0.50	0.50	0.53	0.51
	0.5	6%	0.04	-0.03	-0.02	0.07	4%	0.47	0.51	0.51	0.50
	1	6%	0.14	0.00	0.03	0.10	6%	0.54	0.49	0.50	0.49

Table 11. Median of causal estimates in different quartiles, and proportion of datasets in which the homogeneity test was rejected for Scenario A2 (fixed causal effect of β_1 =0.5) with positive confounding and α_1 =0.1
Proportion homogeneity rejected represents the proportion of simulated datasets where the null hypothesis of homogeneity is rejected

		Positive confounding (α_2 and β_2= 0.8)									
		Stratifying on collider, C					Stratifying on residual collider, C_0				
μ_1	μ_2	Proportion homogeneity rejected (%)	Median estimates Q1	Median estimates Q2	Median estimates Q3	Median estimates Q4	Proportion homogeneity rejected (%)	Median estimates Q1	Median estimates Q2	Median estimates Q3	Median estimates Q4
-1	-1	48%	-0.39	-0.07	0.14	0.58	95%	-0.46	0.19	0.64	1.27
	-0.5	30%	-0.33	-0.06	0.00	0.44	88%	-0.36	0.24	0.59	1.15
	0	19%	-0.25	-0.08	0.00	0.38	78%	-0.25	0.22	0.55	1.09
	0.5	15%	-0.11	0.05	0.14	0.46	61%	-0.19	0.24	0.56	0.99
	1	16%	0.09	0.32	0.44	0.63	40%	-0.09	0.28	0.54	0.88
-0.5	-1	36%	-0.11	0.15	0.32	0.71	68%	-0.07	0.34	0.64	1.08
	-0.5	19%	-0.07	0.14	0.30	0.58	48%	-0.01	0.38	0.64	0.94
	0	14%	0.08	0.23	0.36	0.57	25%	0.11	0.41	0.59	0.87
	0.5	11%	0.22	0.46	0.55	0.76	16%	0.16	0.43	0.56	0.83
	1	16%	0.35	0.58	0.77	0.98	16%	0.19	0.40	0.57	0.81
0	-1	24%	0.24	0.49	0.66	0.95	24%	0.24	0.51	0.70	0.94
	-0.5	14%	0.33	0.52	0.67	0.90	15%	0.33	0.50	0.66	0.89
	0	13%	0.34	0.52	0.66	0.86	13%	0.35	0.53	0.66	0.87
	0.5	13%	0.33	0.51	0.69	0.88	13%	0.33	0.52	0.69	0.89
	1	25%	0.24	0.51	0.69	0.98	26%	0.25	0.52	0.70	0.99
0.5	-1	18%	0.45	0.69	0.87	1.07	18%	0.38	0.59	0.77	1.01
	-0.5	15%	0.34	0.50	0.64	0.88	18%	0.37	0.61	0.79	1.03
	0	14%	0.17	0.30	0.41	0.68	26%	0.30	0.61	0.77	1.07
	0.5	19%	0.05	0.22	0.37	0.72	45%	0.20	0.57	0.84	1.17
	1	34%	0.00	0.25	0.41	0.81	63%	0.13	0.56	0.83	1.27
1	-1	16%	0.25	0.46	0.55	0.88	40%	0.26	0.68	0.90	1.33
	-0.5	12%	0.03	0.16	0.30	0.58	53%	0.19	0.65	0.97	1.37
	0	18%	-0.10	-0.02	0.11	0.49	70%	0.13	0.63	0.97	1.46
	0.5	23%	-0.16	-0.02	0.11	0.60	83%	0.04	0.59	0.98	1.54
	1	35%	-0.23	0.01	0.20	0.71	94%	-0.07	0.55	1.04	1.65

Table 12. Median of causal estimates in different quartiles, and proportion of datasets in which the homogeneity test was rejected for Scenario A3 (varying causal effect) with positive confounding and α_1 =0.1
Proportion homogeneity rejected represents the proportion of simulated datasets where the null hypothesis of homogeneity is rejected

6.6.3. Additional scenarios

In Scenarios B1, B2 and B3, where the collider was a function of both the risk factor and outcome, similar results were observed, with collider bias evident when conditioning on the collider (Table F10) and when stratifying on the collider (Table F11). Collider bias in Scenarios B1 and B2 was greater compared with Scenarios A1 and A2 where the collider was a function of the risk factor only. Similarly, bias was not observed when stratifying on the residual collider (Table F11). For Scenario B3, the power of the homogeneity test was lower in comparison to Scenario A3 (Table F11), as the dependence of effect heterogeneity on the collider was weaker; however, heterogeneity was detected more often when stratifying on the residual collider than on the collider.

For Scenarios C1, C2 and C3, where the outcome was binary, again similar results were observed, with collider bias evident when conditioning on the collider in Scenario C1 (Table F12) and when stratifying on the collider in Scenarios C2 and C3 (Table F13). Bias was smaller than in cases with a continuous outcome, although direct comparison is somewhat unfair as estimates with a binary outcome were obtained from logistic regression and so represent log odds ratios. Estimates when stratifying on the residual collider were slightly attenuated from 0.5 due to the non-collapsibility of the odds ratio (Burgess 2013, 2017). Despite this, in Scenario C2 we observed similar estimates across the different strata of C_0 for each set of parameter values. Similarly, in Scenario C3 we observed that median stratum-specific estimates increased across the four strata when stratifying on either the collider or residual collider. Power to detect heterogeneity was lower compared with Scenario A3 as the stratum-specific estimates are less precise, although again power was consistently higher when stratifying on the residual collider.

6.6.4. Applied example: effect of tobacco smoking on bladder cancer risk across bodyweight strata

Estimates for the causal effect of smoking on bladder cancer in strata of bodyweight and residual bodyweight are shown in Table 13. Estimates represent the odds ratio for bladder cancer per one unit increase in the log odds of being a current smoker. Estimates were positive in all strata, although larger in strata 1 and 2 for both bodyweight and residual bodyweight, and 95% confidence intervals excluded the null in these strata only. Although the homogeneity test was not rejected for either collider variable (p-value = 0.151 and p-value = 0.084 for bodyweight and residual bodyweight, respectively), there was evidence of trend in the stratum-specific estimates for residual bodyweight from meta-regression on the mean value of bodyweight in each stratum (p-value = 0.019). These results suggest that the effect

of smoking on bladder cancer is stronger for subgroups of the population with lower bodyweight.

	Bodyweight Quartile OR [95%CI]				Heterogeneity Test	Trend test
	Q_1	Q_2	Q_3	Q_4	p-value	p-value
Stratifying on bodyweight	**1.59 [1.08; 2.33]**	**1.58 [1.16; 2.14]**	1.13 [0.87; 1.45]	1.11 [0.88; 1.41]	0.151	0.051
Stratifying on residual bodyweight	**1.61 [1.09; 2.37]**	**1.73 [1.28; 2.34]**	1.25 [0.97; 1.62]	1.10 [0.87; 1.39]	0.084	**0.019**

Table 13. Applied example using UK Biobank to investigate the effect of smoking status on bladder cancer risk in different bodyweight strata.
Bodyweight Q_1, Q_2, Q_3, Q_4, represent the four quartiles for both collider and residual collider in which the causal effect of smoking on bladder cancer risk is estimated.
Odds ratios (OR) and 95% confidence intervals (95% CI) for bladder cancer are represent estimates per one unit increase in the log odds of being a current smoker.

6.7. Discussion

In this chapter, we have demonstrated that conditioning or stratifying on a variable that is a collider can have a serious impact on MR estimates. We have introduced a simple approach that constructs a new variable, the residual collider, which is typically highly correlated with the collider, but is independent of the instrument. Estimates obtained from stratification on the residual collider did not suffer from bias in a range of simulation studies. Stratification on the residual collider allows investigators to explore causal estimation in relevant subgroups of the population. We applied our new approach to demonstrate that MR estimates for the effect of smoking on bladder cancer differ within strata of bodyweight, suggesting that the effect of smoking is stronger for subgroups of the population with lower bodyweight.

The approach of stratifying on the residual collider follows the same logic as a previously proposed method for non-linear MR, in which causal estimates are obtained in strata of the population defined by the "residual risk factor" or "IV-free exposure" (Burgess et al. 2014; Staley and Burgess 2017). This variable is defined similarly to the residual collider, except the collider variable is the risk factor itself. This method has been used previously to estimate the causal effect of blood pressure on coronary heart disease risk within strata of blood pressure, resulting in a curve that represents the shape of the causal relationship between the risk factor and the outcome (Malik et al. 2021). This paper extends on that method, showing that the same idea can be used to provide causal estimates stratified on a separate variable even if that variable is a collider. A strength of this method is that its implementation does not depend on the causal structure of the data, in particular the relationships between the collider and other variables in the model.

There are some limitations to this approach. First, while the independence of the residual collider from the instrument is theoretically justified, we demonstrated the validity of our approach through simulation studies. Although we considered a range of different scenarios and parameter values, it is not possible to consider every possible data-generating mechanism by which that a collider could arise. Second, in practice, the relationships between variables are unknown, and so it may be unclear whether a proposed stratifying variable is a collider. However, even if the variable is not a collider, it is unlikely stratification on the residual variable will lead to invalid estimates, suggesting that this approach would be valid for stratifying on variables that are not colliders. This was demonstrated in the simulation study when the effect of the risk factor on the "collider" was zero ($\mu_1 = 0$), and so the stratifying variable was not a collider. One exception is if the stratifying variable is on the causal pathway from the risk factor to the outcome. Stratification on such a variable (a "mediator") will lead to biased estimates even in the proposed approach. Third, the degree of collider bias depends on the strength of the effects of the risk factor and confounder on the collider, and the direction of confounding. Previous work provides an analytical solution to estimate the magnitude of selection bias (Elwert and Segarra 2020). It is possible that collider bias may not be substantial in practice, as observed in the applied example, where estimates were broadly similar when stratifying on bodyweight or residual bodyweight. However, the power to detect heterogeneity in stratum-specific estimates in the simulation study was greater when stratifying on the residual collider, especially when the proportion of variance of the risk factor explained by the instrument was higher. This was also observed in the applied example, where a lower p-value was observed in both the heterogeneity test and the trend test when stratifying on residual bodyweight. Finally, we assumed that the IV assumptions hold; if they do not, estimates will typically be biased. However, several estimation methods that are robust to IV violations are available that allow for consistent estimation under a weaker set of assumptions (Slob and Burgess 2020).

The finding that the effect of smoking on bladder cancer is greater in lower bodyweight subgroups is plausible, because for any given level of cigarette consumption smaller individuals will tend to be exposed to greater concentrations of carcinogens (Luo et al. 2011). An alternative explanation is that the genetic variants could associate more strongly with smoking intensity in individuals of lower bodyweight. However, we would be cautious not to interpret estimates in the higher bodyweight quartiles as implying an absence of a causal effect in heavier individuals; it is possible that the null estimates reflect limited power. Another possible explanation for the results observed is differential survival bias induced by the age of UK Biobank participants. However, as UK Biobank participants were recruited at a

relatively young age (40-65 years), substantial survival bias is unlikely. A limitation of the applied example is overlap between the discovery dataset for the genetic variants, and the dataset used in the MR analysis, which can lead to winner's curse, and the one-sample setting, which can lead to weak instrument bias.

In conclusion, we recommend that researchers performing MR to investigate causal effects in strata of a population defined by a collider stratify on residual values of the collider rather than stratifying on the collider directly.

Chapter 7: Discussion

Chapter 7

In this doctoral thesis, I have extensively reviewed the causal inference methodology, I have applied different statistical approaches to estimate the causal effect of an exposure/treatment on an outcome in both clinical and epidemiological settings under the framework of observational studies and, importantly, I propose further developments to address some of the limitations in the field.

The main limitations of observational studies are the emergence of spurious findings because of the lack of randomization, as well as the presence of confounding bias if proper methodology is not used. In the clinical practice, i.e., in studies that involve intensive care unit (ICU) patients (Chapter 2) and venous thromboembolism patients (Chapter 3), the assigned treatment is usually based on the severity of the disease and randomization is not feasible or unethical. On the other hand, classic epidemiological studies rely on subjects' self-reported information on past exposures allowing to group them as exposed and not-exposed without any randomized intervention. The lack of randomization, in both clinical and epidemiological scenarios, may generate biases in the estimation of the causal effect of interest (Hernán and Robins 2020; Pearl 2010a). This justifies the need to study and develop proper methodology allowing the adjustment for potential confounders to obtain unbiased causal estimates.

The Propensity Score (*PS*) approach is an increasingly popular causal inference method, commonly used in studies involving ICU patients. Using this methods offer several advantages over classic regression models, i.e., logistic regression, and we implemented it to answer two different clinical questions, based on data from an ICU database (Chapter 2).

In the presence of confounding, when the distribution of the baseline covariates differed between treated and untreated patients and also between patients with and without the event, we showed that the two approaches differed in their causal estimates. In particular, we estimated the causal effect of the neuromuscular blockers (the exposure) on delirium events (the outcome) and unlike logistic regression model, a non-significant effect of the exposure was obtained when considering *PS*.

However, in the absence of confounding, when baseline characteristics were balanced between both the exposure and the outcome groups, results from both approaches were similar in terms of point estimation and confidence intervals. Specifically, we studied the effect of the interruption of sedation on mortality where we observed that both PS and classic logistic regression led to analogous results.

The differences between both methodologies (i.e., PS and logistic regression) are related with the frequency of the events and the number of variables used for adjustment. Following Peduzzi et al. (1996), the optimal number of covariates to adjust for in a logistic regression model is related with the number of events observed in the outcome (i.e., 1 covariate per 10 events observed). The same consideration is valid for the Propensity Score, whereas the probability of receiving the treatment is modelled (i.e., 1 covariate per 10 treated patients observed). When the outcome is a rare disease, or it presents low frequency of events, the use of the PS would allow to adjust for more covariates than logistic regression models.

Moreover, similar results in terms of relative odds-ratio and confidence intervals were obtained when comparing *PS* and Randomized Controlled Trials (RCTs). More specifically, we run a systematic review and meta-analysis considering studies involving the assessment of the effect of Venous Thromboembolism (VTE) treatments on patients' mortality (Chapter 3). In most of the scenarios considered, both observational studies and RCT led to similar conclusions.

Therefore, we showed that *PS* is a valid method that can be used by clinicians who are interested in studying the causal effect of a treatment on an outcome when all the potential confounders are measured. However, when confounders are not measured, *PS* is no longer valid, and biased results can be obtained. In such a case,

other methods can be used to estimate a causal effect of interest when unmeasured confounding is present.

The Mendelian Randomization (MR) approach is a novel and appealing causal inference method allowing the estimation of the causal effect of an exposure on an outcome when there are unmeasured confounders.

The advantage of this technique relies on the use of genetic variants as instrumental variables. Here, we showed that different results were obtained by MR and logistic regression models in specific scenarios.

In this framework, we focused on studying the causal effect of T2DM-subtypes (New-Onset Diabetes Mellitus-NODM and Long-Standing Diabetes Mellitus-LSDM) and BMI-obesity on pancreatic cancer risk. We observed a significative association between LSDM and pancreatic cancer using a logistic regression model, i.e., association probably caused by common causes of LSDM and PC. However, using the genetic variants as instrumental variables in a MR model, this association was no longer significant. On the other hand, logistic regression models and MR offered similar results when analysing the causal effect of PC on NODM.

These results offered new insights for pancreatic cancer epidemiology studies as only a few authors have studied the causal effect of diabetes on pancreatic cancer using MR analysis (Carreras-Torres et al. 2017; Lu et al. 2020; Yuan et al. 2020), and none of these studies differentiated between NODM and LSDM. Importantly, our results suggested that although there was no evidence for a causal effect of LSDM on pancreatic cancer risk, pancreatic cancer itself presented a significative causal effect on NODM, which could be considered as an early sign of this disease.

However, despite the advantages that MR presents in terms of confounding bias, other biases may be present. Therefore, we run several sensitivity analyses, such as MR-Egger regression or MR-IVW method, to assess the presence of potential pleiotropy or weak instrument bias. Furthermore, we used a multivariable Mendelian Randomization (MVMR) where both T2DM-subtypes and obesity were considered as exposures and correlated genetic variants were accounted for. With this method pancreatic cancer kept showing a causal effect on NODM, reaching to the same conclusions obtained in the single MR and in the logistic regression model. With that, we showed that MR, including its extensions, can be used to estimate the causal effect of an exposure on an outcome when the no unmeasured confounding assumption does not hold.

There are two main limitations (Bowden et al. 2015) in the context of MR analyses that may generate biased results, and by which interpretations of the results must done carefully. First, due to the low proportion of variation of the exposure explained by the genetic variants, a large sample size is needed in order to obtain precise estimates and to have enough statistical power (Burgess 2014; Burgess and Malarstig 2013; Burgess and Thompson 2011; Davies et al. 2018). In the practice, where the sample size may not be large enough, results could be biased. The second main constraint in MR is related to the IV assumptions, more specific, with the *exclusion restriction* assumption (see Box 3) (Bowden et al. 2015). As VanderWeele et al. (2014) described, there are many scenarios by which this assumption can be violated in the practice, and by which biased results can be obtained. Based on these limitations, cautious interpretations must be done when running MR analyses, considering that causal estimates may be affected by external sources of bias that were not considered in the analysis.

In addition to confounder variables, another type of variables that could also affect the relationship between the exposure and the outcome are the mediators, defined as consequences of the exposure, but causes of the outcome. One approach to model mediation effects in the causal inference field is the Causal Mediation Analysis (CMA). CMA relies on strong assumptions of unmeasured confounders, similarly to *PS*. This technique is not commonly used in cancer epidemiology, but it may offer new insights regarding causal pathways among two exposures and the outcome. We applied, for the first time, CMA to study the mediation effects between T2DM-subtypes and obesity on pancreatic cancer risk. The results obtained suggested that LSDM could act as mediator between obesity and pancreatic cancer and, in the other way around, they showed evidence that obesity can also be considered as a mediator between LSDM and pancreatic cancer.

Because the unmeasured confounding assumption requirement in CMA cannot always hold, we ran a sensitivity analysis considering the *E-value* to evaluate how potential unmeasured confounding would affect the causal estimates of interest, as proposed by Swanson and VanderWeele (2020) The *E-values* obtained suggested that unmeasured confounders were unlikely to explain the effect of the observed association. However, in practice, researchers cannot know whether there are other unmeasured confounders that have not been accounted for but that may introduce bias.

Another important aspect in mediation analysis relates to the requirement of hypotheses regarding the interaction between exposure and mediator, which would

change the definitions of direct, indirect, and total effects (Valeri and VanderWeele 2013). For the development of this thesis, potential interactions between exposure and mediator were excluded.

To address the main limitation of CMA, we proposed a methodological extension, namely Mendelian Randomization in Causal Mediation Analysis (*MRinCMA),* by incorporating two genetic instrumental variables, as in MR, to obtain unbiased estimates of the direct, indirect, and total effects even with potential unmeasured confounding bias.

To assess this new proposal, we considered several simulation scenarios, where we modified variable types, correlation, and effect sizes. We compared the performance of *MRinCMA* with Structural Equation Models (SEM). The results suggested that unbiased estimates can be obtained using this new method. Furthermore, *MRinCMA* outperformed SEM when at least one categorical variable was considered. We applied *MRinCMA* to study the causal effect of obesity on pancreatic cancer risk considering LSDM as a mediator. With this analysis no evidence of causality was found. One potential explanation is the lack of statistical power as a large sample size is required to correctly run *MRinCMA*.

Altogether, *MRinCMA* appears a valid alternative to CMA, where researchers could analyse causal mediation effects without the strong assumptions of confounding bias using genetic variants as instrumental variables for both the exposure and the mediator.

In summary, *PS*, MR, and *MRinCMA* approaches focus mainly in solving and adjusting for potential confounding bias in the estimation of the causal effect. However, other sources of bias may arise in observational studies.

In this regard, collider bias is another form of bias that may occur when we perform a MR analysis stratifying by a collider variable. There is a well-known risk of bias in the stratum-specific estimates if the original collider variable is considered. For that reason, we propose an extension of the MR technique considering the residual collider as an alternative to the original collider. We hypothesised that this proposal could reach unbiased estimates in each stratum. To confirm it, we generated several simulation datasets under different conditions. We showed that residual collider leads to unbiased estimates across all scenarios and presented greater power to detect heterogeneity compared to the original collider. This approach presents two important advantages: first, the interpretation of the obtained results is almost identical between the collider and residual collider and, second, the latter is not

suffering from bias. We applied this approach to study the causal effect of smoking on bladder cancer in different quartiles of bodyweight. We observed that smokers with lower bodyweight presented a higher risk of bladder cancer. This result can be easily translated into clinical practice, justifying specific screening interventions for early prevention of this disease.

However, all these approaches present limitations. As already said, to properly apply *PS* and CMA, all the potential confounder variables must be measured, which is not always feasible using observational studies. Moreover, the true relationships between variables are unknown, and hence it is not clear whether a variable is a collider, or a confounder. Last, large sample sizes are required to run MR analyses, as well as an optimal selection of the genetic variants, to avoid pleiotropy and other genetic biases.

Therefore, I envision further studies to overcome the aforementioned caveats according to the methodology used and the type of data available. Regarding CMA, two main additional studies can be done. First, considering potential interactions between the exposure and mediation, for instance between T2DM-subtypes and obesity. We could extend the *MRinCMA*, as proposed by North et al. (2019), running other simulations when the interaction affects the final causal estimate. Second, by including two or more mediators simultaneously. Recently, some authors developed alternative methodologies to implement CMA with multiple mediators, considering both exposure-mediator and mediator-mediator interactions (Daniel et al. 2015; Steen et al. 2017a; VanderWeele and Vansteelandt 2013). Steen et al. (2017b, 2017a) proposed a flexible mediation analysis to consider multiple mediators. In our epidemiological scenario, to better understand the aetiology of pancreatic cancer, it would be interesting to consider more than one mediator simultaneously, such as pancreatitis, or diabetes treatment.

Further studies can be proposed to study collider bias in MR. Here, we focused on a continuous collider, but in practice, the collider variable of interest may be binary or categorical. For that reason, we could study whether our residual collider approach in MR would be applicable to other variable types. Furthermore, since selection bias is a form of collider bias, (Gkatzionis and Burgess 2019; Munafò et al. 2018), we could study how to implement this residual collider technique in a selection bias framework.

In clinical-epidemiological scenarios, there is an increased use of *omics* data, which refers to the collection of large biomarkers datasets that characterize biological features, including genomics, transcriptomics, proteomics, among others (Hasin et al. 2017; López de Maturana et al. 2019). In general, these types of data can be used

to study the diagnosis and prognosis of diseases, and to better characterize the exposome and the genome. The analysis of omics data presents several statistical and computational challenges (López de Maturana et al. 2019; Pineda et al. 2015). However, only a few authors have considered the use of *omics* data in the causal inference framework until now. For instance, Zuber et al. (2020, 2021) proposed a two-sample multivariable MR approach based on Bayesian model averaging (MR-BMA) for dealing with many highly correlated exposures. They applied MR-BMA on metabolites to prioritise likely causal biomarkers for age-related macular degeneration (Zuber et al. 2020). In a similar context combining MR and Bayesian theory, Howey et al. (2020) proposed a Bayesian network extension to complement MR studies. They analysed the causal effect of fatty acid metabolites on body mass index, using the genetic variants as anchors/instrumental variables. In analogous scenarios, Yazdani et al. (2019) proposed a metabolomic causal network to represent the relationship among metabolites, using the principal components from the genomic variation as instrumental variables to identify a metabolite network. They applied this approach to combine SNPs and metabolites to study the causal determinants of atherosclerosis disease. The use of omics data in the pancreatic cancer studies would help to better understand this disease. For instance, it is known that the gut microbiome has an important role in pancreatic cancer (Li et al. 2020; Wei et al. 2019) and it would be interesting to apply MR-BMA as in Zuber et al. (2020) or a Bayesian network as in Howey et al. (2020) with the most relevant microbiome features to study their causal impact on pancreatic cancer. Furthermore, a relationship between methylation and pancreatic cancer has been previously been established (Mishra and Guda 2017; Tan et al. 2009). Another potential application would be creating a methylation-causal network as in Yazdani et al. (2019) by using genetic variants as instrumental variables to study the effect of the methylation on pancreatic cancer.

Altogether, causal inference methodologies are needed to answer clinical and epidemiological interrogations. The further development of these methods is crucial to continue exploring the causal biological mechanisms behind complex diseases. The methodological extensions proposed in this thesis, as well as the biomedical applications considered, further support the wide range of application that causal inference methods offer and my contributions to the field.

Chapter 8: Conclusions

Chapter 8

The main **methodological** conclusions of this thesis are the following.

First, regarding the application of the Propensity Score approach (Chapter 2):

1. In the absence of confounding bias, the Propensity Score and logistic regression approaches presented similar point estimates and confidence intervals.
2. In the presence of confounding bias, the Propensity Score and logistic regression models presented different point estimates and heterogeneous confidence interval values due to its capacity of adjusting for more covariates than logistic regression models for rare outcomes.

Second, regarding the comparison of the Propensity Score approaches versus Randomized Controlled Trials (Chapter 3):

3. Propensity Score can be considered as valid as Randomized Controlled Trials (i.e., the gold-standard for causality), based on an exhaustive systematic review and meta-analysis and the relative odds ratio obtained, to study the effects of VTE treatments on patient mortality.

Third, regarding the Mendelian Randomization and Causal Mediation Analysis (Chapter 4):

4. Mendelian Randomization and logistic regression models differed in their point estimates and confidence interval values when potential unmeasured confounding bias was present.
5. Mendelian Randomization approaches provided valid causal estimates considering strong instruments and accounting for potential pleiotropy effects.
6. Causal Mediation Analysis was an optimal method to study mediation effects for measured confounders, while unclear results were obtained due to the potential presence of unmeasured confounding bias.

Fourth, regarding the new methodology proposed to answer mediation interrogations, *MRinCMA* (Chapter 5):

7. *MRinCMA* and Structural Equation Models led to similar results when continuous variables were considered.
8. *MRinCMA* performed better than Structural Equation Models when categorical variables are considered, regardless of the sample size, the correlation between variables, and the magnitude of the effect.
9. *MRinCMA* led to unbiased estimates for most of the direct, indirect, and total effects regardless the simulation scenarios.

Fifth, regarding the new approach in Mendelian Randomization used to stratify the population (Chapter 6):

10. The residual collider in Mendelian Randomization was a valid alternative to stratifying on the original collider.
11. Using the residual collider, a greater power to detect heterogeneity in stratum-specific estimates was achieved in comparison to using the original collider.

The main **clinical and epidemiological** conclusions are the following.

First, regarding the clinical ICU scenario (Chapter 2):

1. There was evidence of a causal effect of the interruption of sedation on mortality of patients in ICU, using both Propensity Score and logistic regression models.

2. There was no evidence of a causal effect of neuromuscular blockers on delirium events in ICU patients using Propensity Score analysis.

Second, regarding the clinical VTE scenario (Chapter 3):

3. Different mortality results were obtained according to the treatment assignment for venous thromboembolism patients. Future clinical decisions to improve venous thromboembolism survival must take into consideration not only Randomized Controlled Trials, but also results from observational studies.

Third, regarding the epidemiological scenarios (Chapters 4-6):

4. There was no evidence of a causal effect of LSDM on pancreatic cancer using Mendelian Randomization.
5. There was no evidence of a causal effect of NODM on pancreatic cancer using Mendelian Randomization.
6. There was evidence for a causal effect of pancreatic cancer on NODM using Mendelian Randomization.
7. There was no evidence for a causal effect of obesity on pancreatic cancer using Mendelian Randomization.
8. NODM acted as a mediator between being overweight/obese and pancreatic cancer, using Causal Mediation Analysis.
9. Being overweight/obese acted as a mediator between LSDM and pancreatic cancer, suggesting that LSDM may lead to weight gain, using Causal Mediation Analysis.
10. LSDM acted as a mediator between being overweight/obese and pancreatic cancer.
11. Using the *MRinCMA* proposal, there was no evidence of a causal effect of BMI and obesity on pancreatic cancer, considering LSDM as a mediator.
12. Using the *MRinCMA* proposal, there was no evidence of a causal effect of LSDM on pancreatic cancer, considering obesity as a mediator.
13. The new residual collider approach in Mendelian Randomization showed that smoker with lower bodyweight presented a higher risk of bladder cancer.

References

References

Allemani et al. 2018. "Global Surveillance of Trends in Cancer Survival 2000-14 (CONCORD-3): Analysis of Individual Records for 37 513 025 Patients Diagnosed with One of 18 Cancers from 322 Population-Based Registries in 71 Countries." *Lancet* 391:1023–75. doi: 10.1016/S0140-6736(17)33326-3.

Anglemyer et al. 2014. "Healthcare Outcomes Assessed with Observational Study Designs Compared with Those Assessed in Randomized Trials." *Cochrane Database of Systematic Reviews* 2014(4).

Angrist et al. 1996. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association* 91(434):444–55. doi: 10.1080/01621459.1996.10476902.

Austin. 2011. "An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies." *Multivariate Behavioral Research* 46(3):399–424. doi: 10.1080/00273171.2011.568786.

Baron and Kenny. 1986. "The Moderator-Mediator Variable Distinction in Social Psychological Research. Conceptual, Strategic, and Statistical Considerations." *Journal of Personality and Social Psychology* 51(6):1173–82. doi: 10.1037/0022-3514.51.6.1173.

Belope. 2019. "Data Analysis on Inequalities between Men and Women : Stereotypes , Gender Norms and Shared Responsibility in Childcare."

Boef et al. 2015. "Mendelian Randomization Studies in the Elderly." *Epidemiology (Cambridge, Mass.)* 26(2):e15–16. doi: 10.1097/EDE.0000000000000243.

Bowden et al. 2015. "Mendelian Randomization with Invalid Instruments: Effect Estimation and Bias Detection through Egger Regression." *International Journal of Epidemiology* 44(2):512–25. doi: 10.1093/ije/dyv080.

Bowden et al. 2016. "Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator." *Genetic Epidemiology* 40(4):304–14. doi: 10.1002/gepi.21965.

Burgess. 2013. "Identifying the Odds Ratio Estimated by a Two-Stage Instrumental Variable Analysis with a Logistic Regression Model." *Statistics in Medicine* 32(27):4726–47. doi: 10.1002/sim.5871.

Burgess et al. 2013. "Mendelian Randomization Analysis with Multiple Genetic Variants Using Summarized Data." *Genetic Epidemiology* 37(7):658–65. doi: 10.1002/gepi.21758.

Burgess et al. 2014. "Instrumental Variable Analysis with a Nonlinear Exposure-Outcome Relationship." *Epidemiology* 25(6):877–85. doi: 10.1097/EDE.0000000000000161.

Burgess. 2014. "Sample Size and Power Calculations in Mendelian Randomization with a Single Instrumental Variable and a Binary Outcome." *International Journal of Epidemiology* 43(3):922–29. doi: 10.1093/ije/dyu005.

Burgess. et al. 2015. "A Review of Instrumental Variable Estimators for Mendelian Randomization." *Statistical Methods in Medical Research*. doi: 10.1177/0962280215597579.

Burgess et al. 2015. "Mendelian Randomization: Where Are We Now and Where Are We Going?" *International Journal of Epidemiology* 44(2):379–88. doi: 10.1093/ije/dyv108.

Burgess et al. 2015. "Network Mendelian Randomization: Using Genetic Variants as Instrumental Variables to Investigate Mediation in Causal Pathways." *International Journal of Epidemiology* 44(2):484–95. doi: 10.1093/ije/dyu176.

Burgess et al. 2016. "Beyond Mendelian Randomization: How to Interpret Evidence of Shared Genetic Predictors." *Journal of Clinical Epidemiology* 69:208–16. doi: 10.1016/j.jclinepi.2015.08.001.

Burgess et al. 2017. "A Review of Instrumental Variable Estimators for Mendelian Randomization." *Statistical Methods in Medical Research* 26(5):2333–55. doi: 10.1177/0962280215597579.

Burgess. 2017. "Estimating and Contextualizing the Attenuation of Odds Ratios Due to Non Collapsibility." *Communications in Statistics - Theory and Methods* 46(2):786–804. doi: 10.1080/03610926.2015.1006778.

Burgess et al. 2017. "Sensitivity Analyses for Robust Causal Inference from Mendelian Randomization Analyses with Multiple Genetic Variants." *Epidemiology* 28(1):30–42. doi: 10.1097/EDE.0000000000000559.

Burgess and Malarstig. 2013. "Using Mendelian Randomization to Assess and Develop Clinical Interventions: Limitations and Benefits." *Journal of Comparative Effectiveness Research* 2(3):209–12. doi: 10.2217/cer.13.14.

Burgess and Thompson. 2011. "Bias in Causal Estimates from Mendelian Randomization Studies with Weak Instruments." *Statistics in Medicine* 30(11):1312–23. doi: 10.1002/sim.4197.

Burgess and Thompson. 2013. "Use of Allele Scores as Instrumental Variables for Mendelian Randomization." *International Journal of Epidemiology* 42(4):1134–44. doi: 10.1093/ije/dyt093.

Burgess and Thompson. 2015a. *Mendelian Randomization Methods for Using Genetic Variants in Causal Estimation*. Chapman & Hall/CRC.

Burgess and Thompson. 2015b. "Multivariable Mendelian Randomization: The Use of Pleiotropic Genetic Variants to Estimate Causal Effects." *American Journal of Epidemiology* 181(4):251–60. doi: 10.1093/aje/kwu283.

Burgess and Thompson. 2017. "Interpreting Findings from Mendelian Randomization Using the MR-Egger Method." *European Journal of Epidemiology* 32(5):377–89. doi: 10.1007/s10654-017-0255-x.

Canan et al. 2017. "Instrumental Variable Analyses and Selection Bias." *Epidemiology* 28(3):396–98. doi: 10.1097/EDE.0000000000000639.

Carreras-Torres et al. 2017. "The Role of Obesity and Metabolic Factors in Pancreatic Cancer: A Mendelian Randomization Study." *J Natl Cancer Inst* 109(9). doi: 10.1093/jnci/djx012.

Carter et al. 2021. "Mendelian Randomisation for Mediation Analysis: Current Methods and Challenges for Implementation." *European Journal of Epidemiology* 36(5):465–78. doi: 10.1007/s10654-021-00757-1.

Coscia et al. 2019. "Comparison of All-Cause Mortality Following VTE Treatment Between Propensity Score-Adjusted Observational Studies and Matched Randomized Controlled Trials Meta-Epidemiologic Study." *CHEST* 155:689–98. doi: 10.1016/j.chest.2018.10.016.

Coscia Requena et al. 2018. "Analysis of Causality from Observational Studies and Its Application in Clinical Research in Intensive Care Medicine." *Medicina Intensiva*. doi: 10.1016/j.medin.2018.01.002.

Daniel et al. 2015. "Causal Mediation Analysis with Multiple Mediators." *Biometrics* 71(1):1–14. doi: 10.1111/biom.12248.

Davies et al. 2018. "Reading Mendelian Randomisation Studies: A Guide, Glossary, and Checklist for Clinicians." *BMJ (Online)* 362:601. doi: 10.1136/bmj.k601.

Elwert and Segarra. 2020. "Instrumental Variables with Treatment-Induced Selection: Exact Bias Results." *ArXiv*.

Esterhuizen and Thabane. 2016. "Con: Meta-Analysis: Some Key Limitations and Potential Solutions." *Nephrology Dialysis Transplantation* 31(6):882–85.

Faraoni and Schaefer. 2016. "Randomized Controlled Trials vs. Observational Studies: Why Not Just Live Together?" *BMC Anesthesiology* 16(1).

Frölich and Huber. 2017. "Direct and Indirect Treatment Effects–Causal Chains and Mediation Analysis with Instrumental Variables." *Journal of the Royal Statistical Society. Series B: Statistical Methodology* 79(5):1645–66. doi: 10.1111/rssb.12232.

Ganeshkumar and Gopalakrishnan. 2013. "Systematic Reviews and Meta-Analysis: Understanding the Best Evidence in Primary Healthcare." *Journal of Family Medicine and Primary Care* 2(1):9. doi: 10.4103/2249-4863.109934.

Gkatzionis and Burgess. 2019. "Contextualizing Selection Bias in Mendelian Randomization: How Bad Is It Likely to Be?" *International Journal of Epidemiology* 48(3):691–701. doi: 10.1093/ije/dyy202.

Greenland. 2003. "Quantifying Biases in Causal Models: Classical Confounding vs Collider-Stratification Bias." *Epidemiology* 14(3):300–306. doi: 10.1097/01.EDE.0000042804.12056.6C.

Hannan. 2008. "Randomized Clinical Trials and Observational Studies. Guidelines for Assessing Respective Strengths and Limitations." *JACC: Cardiovascular Interventions* 1(3):211–17. doi: 10.1016/j.jcin.2008.01.008.

Hasin et al. 2017. "Multi-Omics Approaches to Disease." *Genome Biology* 18(1).

Haycock et al. 2016. "Best (but Oft-Forgotten) Practices: The Design, Analysis, and Interpretation of Mendelian Randomization Studies." *American Journal of Clinical Nutrition* 103(4):965–78. doi: 10.3945/ajcn.115.118216.

Hernán et al. 2004. "A Structural Approach to Selection Bias." *Epidemiology* 15(5):615–25. doi: 10.1097/01.ede.0000135174.63482.43.

Hernán and Robins. 2006. "Instruments for Causal Inference: An Epidemiologist's Dream?" *Epidemiology*. doi: 10.1097/01.ede.0000222409.00878.37.

Hernán and Robins. 2020. *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC.

Higgins et al. 2003. "Measuring Inconsistency in Meta-Analyses." *British Medical Journal* 327(7414):557–60.

Howey et al. 2020. *Bayesian Network Analysis Incorporating Genetic Anchors Complements Conventional Mendelian Randomization Approaches for Exploratory Analysis of Causal Relationships in Complex Data*. Vol. 16.

Hu and Mustillo. 2016. "Recent Development of Propensity Score Methods in Observational Studies : Multi- Categorical Treatment, Causal Mediation, and Heterogeneity Background: Propensity Score Methods in the Counterfactual Framework." *Current Sociology Imbens and Rubin Imbens and Wooldridge* 64(1):60–82. doi: 10.1177/0011392115589599.

Huang et al. 2021. "Worldwide Burden of, Risk Factors for, and Trends in Pancreatic Cancer." *Gastroenterology* 160(3):744–54. doi: 10.1053/j.gastro.2020.10.007.

Hudgens and Halloran. 2008. "Toward Causal Inference with Interference." *Journal of the American Statistical Association* 103(482):832–42. doi: 10.1198/016214508000000292.

Hughes et al. 2019. "Selection Bias When Estimating Average Treatment Effects Using One-Sample Instrumental Variable Analysis." *Epidemiology* 30(3):350–57. doi: 10.1097/EDE.0000000000000972.

Imai et al. 2010. "A General Approach to Causal Mediation Analysis." doi: 10.1037/a0020761.

Imai et al. 2011. "Commentary: Using Potential Outcomes to Understand Causal Mediation Analysis." *Multivariate Behavioral Research* 46(5):861–73. doi: 10.1080/00273171.2011.606743.

Imbens and Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica* 62(2):467–75.

Imbens and Rubin. 2015. *Causal Inference for Statistics, Social, and Biomedical Sciences*. Cambridge University Press.

Kato et al. 2020. "Propensity Score-Weighted Analysis of Chemotherapy after PD-1 Inhibitors versus Chemotherapy Alone in Patients with Non-Small Cell Lung Cancer (WJOG10217L)." *Journal for ImmunoTherapy of Cancer* 8(1):350. doi: 10.1136/jitc-2019-000350.

Kent et al. 2017. "Body Mass Index and Healthcare Costs: A Systematic Literature Review of Individual Participant Data Studies." *Obesity Reviews* 18(8):869–79.

Larsson et al. 2020. "Smoking, Alcohol Consumption, and Cancer: A Mendelian Randomisation Study in UK Biobank and International Genetic Consortia Participants." *PLoS Medicine* 17(7):1–14. doi: 10.1371/journal.pmed.1003178.

Lawlor et al. 2008. "Mendelian Randomization: Using Genes as Instruments for Making Causal Inferences in Epidemiology." *Stat Med2* 27(8):1133–63. doi: 10.1002/sim.3034.

Li. 2016. "The Performance of ML, DWLS, and ULS Estimation with Robust Corrections in Structural Equation Models with Ordinal Variables." *Psychological Methods* 21(3):369–87. doi: 10.1037/met0000093.

Li et al. 2020. "Gut Microbiota: Its Potential Roles in Pancreatic Cancer." *Frontiers in Cellular and Infection Microbiology* 10:564.

Liu et al. 2019. "Association Studies of up to 1.2 Million Individuals Yield New Insights into the Genetic Etiology of Tobacco and Alcohol Use." *Nature Genetics* 51(2):237–44. doi: 10.1038/s41588-018-0307-5.

López de Maturana et al. 2019. "Challenges in the Integration of Omics and Non-Omics Data." *Genes* 10(3).

López de Maturana et al. 2021. "A Multilayered Post-GWAS Assessment on Genetic

Susceptibility to Pancreatic Cancer." *Genome Medicine* 13(1). doi: 10.1186/s13073-020-00816-4.
Lu et al. 2020. "Mendelian Randomisation Study of the Effects of Known and Putative Risk Factors on Pancreatic Cancer." *Journal of Medical Genetics* 57(12):820–28. doi: 10.1136/jmedgenet-2019-106200.
Luo et al. 2011. "Interaction between Smoking and Obesity and the Risk of Developing Breast Cancer among Postmenopausal Women." *American Journal of Epidemiology* 174(8):919–28. doi: 10.1093/aje/kwr192.
Maisonneuve and Lowenfels. 2015. "Risk Factors for Pancreatic Cancer: A Summary Review of Meta-Analytical Studies." *International Journal of Epidemiology* 44(1):186–98. doi: 10.1093/ije/dyu240.
Malats and Real. 2015. "Epidemiology of Bladder Cancer." *Hematology/Oncology Clinics of North America* 29(2):177–89.
Malik et al. 2021. "Relationship between Blood Pressure and Incident Cardiovascular Disease: Linear and Nonlinear Mendelian Randomization Analyses." *Hypertension* (June):2004–13. doi: 10.1161/HYPERTENSIONAHA.120.16534.
Martinussen et al. 2017. "Instrumental Variables Estimation of Exposure Effects on a Time-to-Event Endpoint Using Structural Cumulative Survival Models." *Biometrics* 73(4):1140–49. doi: 10.1111/biom.12699.
Mishra and Guda. 2017. "Genome-Wide DNA Methylation Analysis Reveals Molecular Subtypes of Pancreatic Cancer." *Oncotarget* 8(17):28990–12. doi: 10.18632/oncotarget.15993.
Molina-Montes et al. 2021. "Deciphering the Complex Interplay between Pancreatic Cancer, Diabetes Mellitus Subtypes and Obesity/BMI through Causal Inference and Mediation Analyses." *Gut* 70:319–29. doi: 10.1136/gutjnl-2019-319990.
Morris et al. 2019. "Using Simulation Studies to Evaluate Statistical Methods." *Statistics in Medicine* 38(11):2074–2102. doi: 10.1002/sim.8086.
Munafò et al. 2009. "Smoking Status and Body Mass Index: A Longitudinal Study." *Nicotine and Tobacco Research* 11(6):765–71. doi: 10.1093/ntr/ntp062.
Munafò et al. 2018. "Collider Scope: When Selection Bias Can Substantially Influence Observed Associations." *International Journal of Epidemiology* 47(1):226–35. doi: 10.1093/ije/dyx206.
Murad et al. 2016. "New Evidence Pyramid." *Evidence-Based Medicine* 21(4):125–27.
Muthén. 1984. "A General Structural Equation Model with Dichotomous, Ordered Categorical, and Continuous Latent Variable Indicators." *Psychometrika* 49(1):115–32. doi: 10.1007/BF02294210.
Muthén. 2011. "Applications of Causally Defined Direct and Indirect Effects in Mediation Analysis Using SEM in Mplus." *Psychological Bulletin* 107(2):238–46.
Muthén and Asparouhov. 2015. "Causal Effects in Mediation Modeling: An Introduction With Applications to Latent Variables." *Structural Equation Modeling* 22(1):12–23. doi: 10.1080/10705511.2014.935843.
North et al. 2019. "Using Genetic Instruments to Estimate Interactions in Mendelian Randomization Studies." *Epidemiology* 30(6):E33–35.
Olsson et al. 2000. "The Performance of ML, GLS, and WLS Estimation in Structural Equation Modeling under Conditions of Misspecification and Nonnormality." *Structural Equation Modeling* 7(4):557–95. doi: 10.1207/S15328007SEM0704_3.
Palmer et al. 2011. "Instrumental Variable Estimation of Causal Risk Ratios and Causal Odds Ratios in Mendelian Randomization Analyses." *American Journal of Epidemiology* 173(12):1392–1403. doi: 10.1093/aje/kwr026.
Palmer et al. 2017. "Correcting the Standard Errors of 2-Stage Residual Inclusion Estimators for Mendelian Randomization Studies." *American Journal of Epidemiology* 186(9):1104–14. doi: 10.1093/aje/kwx175.
Paternoster et al. 2017. "Genetic Epidemiology and Mendelian Randomization for Informing Disease Therapeutics: Conceptual and Methodological Challenges." *PLoS Genetics* 13(10):e1006944. doi: 10.1371/journal.pgen.1006944.
Pearl. 1995. "Causal Diagrams for Empirical Research." *Biometrika* 82(4):669–88. doi: 10.1093/biomet/82.4.669.
Pearl. 2001. "Direct and Indirect Effects." *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence* 411–20.
Pearl. 2010a. "An Introduction to Causal Inference." *International Journal of Biostatistics* 6(2). doi: 10.1097/EDE.0000000000001164.
Pearl. 2010b. "On the Consistency Rule in Causal Inference: Axiom, Definition, Assumption, or Theorem?" *Epidemiology* 21(6):872–75. doi: 10.1097/EDE.0b013e3181f5d3fd.
Pearl. 2012a. "The Causal Foundations of Structural Equation Modeling." *Handbook of Structural Equation Modeling* (June):68–91.

Pearl. 2012b. "The Causal Mediation Formula-A Guide to the Assessment of Pathways and Mechanisms." *Prevention Science* 13(4):426–36. doi: 10.1007/s11121-011-0270-1.
Peduzzi et al. 1996. "A Simulation Study of the Number of Events per Variable in Logistic Regression Analysis." *Journal of Clinical Epidemiology* 49(12):1373–79. doi: 10.1016/S0895-4356(96)00236-3.
Piirtola et al. 2018. "Association of Current and Former Smoking with Body Mass Index: A Study of Smoking Discordant Twin Pairs from 21 Twin Cohorts." *PLoS ONE* 13(7). doi: 10.1371/journal.pone.0200140.
Pineda et al. 2015. "Integration Analysis of Three Omics Data Using Penalized Regression Methods: An Application to Bladder Cancer." *PLoS Genetics* 11(12):1–22. doi: 10.1371/journal.pgen.1005689.
Rahib et al. 2014. "Projecting Cancer Incidence and Deaths to 2030: The Unexpected Burden of Thyroid, Liver, and Pancreas Cancers in the United States." *Cancer Research* 74(11):2913–21.
Rees et al. 2017. "Extending the MR-Egger Method for Multivariable Mendelian Randomization to Correct for Both Measured and Unmeasured Pleiotropy." *Statistics in Medicine* 36(29):4705–18. doi: 10.1002/sim.7492.
Relton and Davey Smith. 2012. "Two-Step Epigenetic Mendelian Randomization: A Strategy for Establishing the Causal Role of Epigenetic Processes in Pathways to Disease." *International Journal of Epidemiology* 41(1):161–76. doi: 10.1093/ije/dyr233.
Ren et al. 2015. "Randomized Controlled Trials of Induction Treatment and Surgery versus Combined Chemotherapy and Radiotherapy in Stages IIIA-N2 NSCLC: A Systematic Review and Meta-Analysis." *Journal of Thoracic Disease* 7(8):1414–22. doi: 10.3978/j.issn.2072-1439.2015.08.14.
Del Rey Calero et al. 1996. *Fundamentos de Epidemiología Clínica*.
Robins and Greenland. 1992. "Identifiability and Exchangeability for Direct and Indirect Effects." *Epidemiology* 3(2):143–55.
Rosenbaum and Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." Pp. 41–55 in *Biometrika*. Vol. 70.
Rosenbaum and Rubin. 1984. "Reducing Bias in Observational Studies Using Subclassification on the Propensity Score." *Journal of the American Statistical Association* 79(387):516–24. doi: 10.1080/01621459.1984.10478078.
Rosner. 2012. "Evidence-Based Medicine: Revisiting the Pyramid of Priorities." *Journal of Bodywork and Movement Therapies* 16(1):42–49. doi: 10.1016/j.jbmt.2011.05.003.
Rosseel. 2012. "Lavaan: An R Package for Structural Equation Modeling." *Journal of Statistical Software* 48(2). doi: 10.18637/jss.v048.i02.
Rubin. 1974. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology* 66(5):688–701.
Sanderson. 2021. "Multivariable Mendelian Randomization and Mediation." *Cold Spring Harbor Perspectives in Medicine* 11(2):1–12. doi: 10.1101/cshperspect.a038984.
Schad et al. 2018. "Overall Survival of Stage IV Non-Small Cell Lung Cancer Patients Treated with Viscum Album L. In Addition to Chemotherapy, a Real-World Observational Multicenter Analysis." *PLoS ONE* 13(8). doi: 10.1371/journal.pone.0203058.
Sheehan and Didelez. 2020. "Epidemiology, Genetic Epidemiology and Mendelian Randomisation: More Need than Ever to Attend to Detail." *Human Genetics* 139(1):121–36. doi: 10.1007/s00439-019-02027-3.
Slob and Burgess. 2020. "A Comparison of Robust Mendelian Randomization Methods Using Summary Data." *Genetic Epidemiology* 44(4):313–29. doi: 10.1002/gepi.22295.
Small et al. 2017. "Instrumental Variable Estimation with a Stochastic Monotonicity Assumption." *Statistical Science* 32(4):561–79. doi: 10.1214/17-STS623.
Smit et al. 2019. "Survival Bias in Mendelian Randomization Studies: A Threat to Causal Inference." *Epidemiology (Cambridge, Mass.)* 30(6):813–16. doi: 10.1097/EDE.0000000000001072.
Smith and Ebrahim. 2003. "'Mendelian Randomization': Can Genetic Epidemiology Contribute to Understanding Environmental Determinants of Disease?" *International Journal of Epidemiology* 32(1):1–22.
Smith and Ebrahim. 2005. "What Can Mendelian Randomisation Tell Us about Modifiable Behavioural and Environmental Exposures?" *British Medical Journal* 330(7499):1076–79.
Staley and Burgess. 2017. "Semiparametric Methods for Estimation of a Nonlinear Exposure-Outcome Relationship Using Instrumental Variables with Application to Mendelian Randomization." *Genetic Epidemiology* 41(4):341–52. doi: 10.1002/gepi.22041.
De Stavola et al. 2015. "Practice of Epidemiology Mediation Analysis With Intermediate Confounding: Structural Equation Modeling Viewed Through the Causal Inference Lens." *American Journal of Epidemiology* 181(1):64–80. doi: 10.1093/aje/kwu239.

Steen et al. 2017a. "Flexible Mediation Analysis with Multiple Mediators." *American Journal of Epidemiology* 186(2):184–93. doi: 10.1093/aje/kwx051.
Steen et al. 2017b. "Medflex: An R Package for Flexible Mediation Analysis Using Natural Effect Models." *Journal of Statistical Software* 76(1). doi: 10.18637/jss.v076.i11.
Sudlow et al. 2015. "UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age." *PLoS Medicine* 12(3):1–10. doi: 10.1371/journal.pmed.1001779.
Swanson et al. 2015. "Definition and Evaluation of the Monotonicity Condition for Preference-Based Instruments." *Epidemiology* 26(3):414–20. doi: 10.1097/eDe.0000000000000279.
Swanson. 2019. "A Practical Guide to Selection Bias in Instrumental Variable Analyses." *Epidemiology* 30(3):345–49.
Swanson and Hernan. 2018. "The Challenging Interpretation of Instrumental Variable Estimates under Monotonicity." *International Journal of Epidemiology* 47(4):1289–97. doi: 10.1093/ije/dyx038.
Swanson and Hernán. 2013. "Commentary: How to Report Instrumental Variable Analyses (Suggestions Welcome)." *Epidemiology (Cambridge, Mass.)* 24(3):370–74. doi: 10.1097/EDE.0B013E31828D0590.
Swanson and VanderWeele. 2020. "E-Values for Mendelian Randomization." *Epidemiology* 31(3):E23–24. doi: 10.1097/EDE.0000000000001164.
Tan et al. 2009. "Characterizing DNA Methylation Patterns in Pancreatic Cancer Genome." *Molecular Oncology* 3(5–6):425–38. doi: 10.1016/j.molonc.2009.03.004.
Taylor et al. 2019. "The Effect of Body Mass Index on Smoking Behaviour and Nicotine Metabolism: A Mendelian Randomization Study." *Human Molecular Genetics* 28(8):1322–30. doi: 10.1093/hmg/ddy434.
Tchetgen et al. 2015. "Instrumental Variable Estimation in a Survival Context." *Epidemiology (Cambridge, Mass.)* 26(3):402–10. doi: 10.1097/EDE.0000000000000262.
Tchetgen and Vanderweele. 2012. "On Causal Inference in the Presence of Interference." *Statistical Methods in Medical Research* 21(1):55–75. doi: 10.1177/0962280210386779.
Terza et al. 2008. "Two-Stage Residual Inclusion Estimation: Addressing Endogeneity in Health Econometric Modeling." *Journal of Health Economics* 27(3):531–43. doi: 10.1016/j.jhealeco.2007.09.009.
Thompson and Higgins. 2002. "How Should Meta-Regression Analyses Be Undertaken and Interpreted?" *Statistics in Medicine* 21(11):1559–73. doi: 10.1002/sim.1187.
Uddin et al. 2015. "Instrumental Variable Analysis in Epidemiologic Studies: An Overview of the Estimation Methods." *Pharmaceutica Analytica Acta* 06(04). doi: 10.4172/2153-2435.1000353.
Valeri and VanderWeele. 2013. "Mediation Analysis Allowing for Exposure-Mediator Interactions and Causal Interpretation: Theoretical Assumptions and Implementation with SAS and SPSS Macros." *Psychological Methods* 18(2):137–50. doi: 10.1037/a0031034.
VanderWeele. 2009. "Concerning the Consistency Assumption in Causal Inference." *Epidemiology* 20(6):880–83. doi: 10.1097/EDE.0b013e3181bd5638.
VanderWeele et al. 2014. "Methodological Challenges in Mendelian Randomization." *Epidemiology* 25(3):427–35. doi: 10.1097/EDE.0000000000000081.
VanderWeele. 2016. "Mediation Analysis: A Practitioner's Guide." *Annual Review of Public Health* 37:17–32.
VanderWeele and Vansteelandt. 2009. "Conceptual Issues Concerning Mediation, Interventions and Composition." *Statistics and Its Interface* 2(4):457–68. doi: 10.4310/sii.2009.v2.n4.a7.
VanderWeele and Vansteelandt. 2013. "Mediation Analysis with Multiple Mediators." *Epidemiologic Methods* 2(1):95–115. doi: 10.1515/em-2012-0010.
Wei et al. 2019. "The Microbiota and Microbiome in Pancreatic Cancer: More Influential than Expected." *Molecular Cancer* 18(1).
Yazdani et al. 2019. "Genome Analysis and Pleiotropy Assessment Using Causal Networks with Loss of Function Mutation and Metabolomics." *BMC Genomics* 20(1). doi: 10.1186/s12864-019-5772-4.
Yuan et al. 2020. "Is Type 2 Diabetes Causally Associated with Cancer Risk? Evidence from a Two-Sample Mendelian Randomization Study." *Diabetes* 69(7):1588–96. doi: 10.2337/db20-0084.
Zenke et al. 2021. "Effect of Second-Generation vs Third-Generation Chemotherapy Regimens with Thoracic Radiotherapy on Unresectable Stage III Non-Small-Cell Lung Cancer: 10-Year Follow-up of a WJTOG0105 Phase 3 Randomized Clinical Trial." *JAMA Oncology* 7(6):904–9. doi: 10.1001/jamaoncol.2021.0113.

Zheng et al. 2017. "Recent Developments in Mendelian Randomization Studies." *Current Epidemiology Reports* 4(4):330–45. doi: 10.1007/s40471-017-0128-6.

Zuber et al. 2020. "Selecting Likely Causal Risk Factors from High-Throughput Experiments Using Multivariable Mendelian Randomization." *Nature Communications* 11(1). doi: 10.1038/s41467-019-13870-3.

Zuber et al. 2021. "High-Throughput Multivariable Mendelian Randomization Analysis Prioritizes Apolipoprotein B as Key Lipid Risk Factor for Coronary Artery Disease." *International Journal of Epidemiology* 50(3):893–901. doi: 10.1093/IJE/DYAA216.

Appendices

Appendix A. Supplementary Material Chapter 4

Index of Supplementary Tables, Chapter 4

Index of Supplementary Figures, Chapter 4

Supplementary Material.

PanGenEU centres and investigators

Spanish National Cancer Research Centre (CNIO), Madrid, Spain: Núria Malats[1], Francisco X Real[1], Evangelina López de Maturana, Paulina Gómez-Rubio, Esther Molina-Montes, Lola Alonso, Mirari Márquez, Roger Milne, Ana Alfaro, Tania Lobato, Lidia Estudillo.
Verona University, Italy: Rita Lawlor[1], Aldo Scarpa, Stefania Beghelli.
National Cancer Registry Ireland, Cork, Ireland: Linda Sharp[1], Damian O'Driscoll.
Hospital Madrid-Norte-Sanchinarro, Madrid, Spain: Manuel Hidalgo[1], Jesús Rodríguez Pascual.
Hospital Ramon y Cajal, Madrid, Spain: Alfredo Carrato[1], Carmen Guillén-Ponce, Mercedes Rodríguez-Garrote, Federico Longo-Muñoz, Reyes Ferreiro, Vanessa Pachón, M Ángeles Vaz.

Hospital del Mar, Barcelona, Spain: Mar Iglesias[1], Cristina Álvarez-Urturi, Xavier Bessa, Felipe Bory, Lucas Ilzarbe, Lucía Márquez, Ignasi Poves✝, Fernando Burdío, Luis Grande, Javier Gimeno.
Hospital Vall d´Hebron, Barcelona, Spain: Xavier Molero[1], Luisa Guarner✝, Joaquin Balcells.
Technical University of Munich, Germany: Christoph Michalski[1], Jörg Kleeff, Bo Kong.
Karolinska Institute, Stockholm, Sweden: Matthias Löhr[1], Jiaqui Huang, Weimin Ye, Jingru Yu.

Hospital 12 de Octubre, Madrid, Spain: José Perea[1], Pablo Peláez.
Hospital de la Santa Creu i Sant Pau, Barcelona, Spain: Antoni Farré[1], Josefina Mora, Marta Martín, Vicenç Artigas, Carlos Guarner, Francesc J Sancho, Mar Concepción, Teresa Ramón y Cajal.
The Royal Liverpool University Hospital, UK: William Greenhalf[1], Eithne Costello.
Queen's University Belfast, UK: Michael O'Rorke[1], Liam Murray✝, Marie Cantwell.
Laboratorio de Genética Molecular, Hospital General Universitario de Elche, Spain: Víctor M Barberá[1], Javier Gallego.
Instituto Universitario de Oncología del Principado de Asturias, Oviedo, Spain: Adonina Tardón[1], Luis Barneo.
Hospital Clínico Universitario de Santiago de Compostela, Spain: Enrique Domínguez Muñoz[1], Antonio Lozano, Maria Luaces.
Hospital Clínico Universitario de Salamanca, Spain: Luís Muñoz-Bellvís[1], J.M. Sayagués, M.L. Gutíerrrez Troncoso, A. Orfao de Matos.
University of Marburg, Department of Gastroenterology, Phillips University of Marburg, Germany: Thomas Gress[1], Malte Buchholz, Albrecht Neesse.
Queen Mary University of London, UK: Tatjana Crnogorac-Jurcevic[1], Hemant M Kocher, Satyajit Bhattacharya, Ajit T Abraham, Darren Ennis, Thomas Dowe, Tomasz Radon
Scientific advisors of the PanGenEU Study: Debra T Silverman (NCI, USA) and Douglas Easton (U. of Cambridge, UK)

[1] Principal Investigator in each centre

Supplementary Methods

Study population: PanGenEU (the European Study into Digestive Illnesses and Genetics) is a mostly hospital-based case-control study of PC conducted in six European countries (Spain, Germany, Ireland, United Kingdom, Italy and Sweden) and 28 centers, designed to evaluate environmental and genetic factors associated with PC. Recruitment of PC cases and corresponding controls matched by region, sex and age (± 10 years) took place from 2007 to 2014 in all participating centers, except in those from Italy were only cases were ascertained. Inclusion criteria were cases diagnosed or suspected of having PC, who had lived in one of the study areas and aged older than 18 years. All medical records were reviewed to ensure the PC diagnosis for study entry. Participants incapable of participating in the study due to impairment of physical ability were excluded. Response rates varied by center and were on average 76% among cases and 85% among controls.

Data collection: A standardized epidemiological questionnaire including self-reported socio-demographic and anthropometric data (location of body fat, height and weight at different ages: age 20 and 50 years, 2 years before recruitment and at PC diagnosis), the likely fat accumulation zone (abdominal, hips, all equally, no extra weight gain), family history of cancer including PC, medical history (e.g., chronic pancreatitis, diabetes and others) including regular use of specific medication, and lifestyle behaviors (e.g., smoking and alcohol habits) was administered by trained personnel in a face-to-face interview. This information was used as input to generate other variables such as body mass index (BMI: weight in kg / height in m^2: <25, 25-30, ≥30 kg/m^2) at different ages (20, 50 and two years before recruitment). Weight gain (> 5 or 10 kg) between 20 and 50 years, and weight loss from age 50 until two years before recruitment, was also derived (younger than 50 years, yes weight gain/loss, no weight gain/loss).

T2DM biomarker assays: Non-fasting erythrocyte and serum samples collected at subject recruitment and stored at -80ºC from 509 PC cases and 413 controls of the Spanish PanGenEU study were analysed blinded to the disease status. All individuals had epidemiological information; 356 cases and 298 controls also participated in the genetic study. Glycated haemoglobin or Hb1Ac (as percentage of haemoglobin and mmol/mol) was measured with an automated HPLC analyzer (Menarini Diagnostics, Spain) at the Hospital 12 de Octubre, Madrid (Spain). Mean intra-batch and inter-batch coefficients of variations were 0.42% and 8.46%, respectively. Diabetes status based on Hb1Ac data was established for values above 6.5%. Undiagnosed T2DM, most likely NODM, was identified on this basis. Other predefined levels of Hb1Ac were considered to distinguish between prediabetes (≥6% and <6.5%) and non-diabetes (<6%). Furthermore, undiagnosed or uncontrolled T2DM (Hb1Ac≥6.5%), or HbA1c levels <6.5% but self-reported T2DM diagnosis, i.e., controlled T2DM patients, were considered in separate categories. C-peptide was measured at University Hospital Giessen and Marburg using a Cobas e411 (Roche Diagnostics, Mannheim, Germany) by means of Electro-

chemiluminescence immune assay. Coefficients of variations were <5%. Type 3c-like diabetes was defined as NODM with C-peptide levels below the median (4.2 µg/L in controls).

SNP selection and genotyping: Consistent quality SNP data was available for 1,162 cases and 540 controls who provided blood samples. DNA samples were genotyped on the Infinium OncoArray-500K at the CEGEN (Spanish National Cancer Research Centre, CNIO). The genotype data was filtered for call rate, relatedness, European ancestry <80% and sex chromosome abnormalities. Overall, 451,883 SNPs passed these quality filters and underwent imputation of missing genotypes using IMPUTE v2. The control group was enlarged with 212 controls participating in two Spanish bladder cancer case-control studies (EPICURO and ISBlaC), with analogous characteristics to the source population (Spanish PC cases; 44% females and mean age=64.7 years) and with genotype and epidemiological data available. Similar protocols for data collection and genotyping were used in all studies. Genotype distributions of each SNP and deviation from Hardy-Weinberg equilibrium were assessed separately in each of the geographical areas. Principal components to control for population stratification were calculated with the *prcomp* function in *R*.

Imputation: Missing data (9.8% in the dataset containing 63 variables with predictors to improve the imputation performance), assumed to be at random, was substituted by the *random forest* (RF) imputation algorithm. RF tress (n=100) trained on the observed values of the data set predicted the missing values of the data. The out-of-bag error (OOB) was considered as a measure of the imputation error. A further test of imputation performance consisted of comparing observed imputed values *versus* the expected values (% of concordance) in a test set resembling the pattern of missingness of the original data. Mean OOB error (0.05) and concordance estimates (92.5%) demonstrated good imputation performance (Supplementary Table 2).

Statistical analysis for the observational association study: There were 2,018 PC cases and 1,540 controls available for assessing the observational association between T2DM and PC risk (Supplementary Figure 1A). Descriptive statistics by case-control status were performed, evaluating differences between the groups via Pearson chi-square and Student's t-test or Mann-Whitney test, where appropriate. Multivariate unconditional logistic regression was applied to evaluate the association between T2DM and PC risk by Odds Ratios (ORs) and 95% Confidence Intervals (CIs). The influence of smoking, obesity (BMI variables), alcohol status, asthma and/or allergies, educational level, and family history of PC, was evaluated in age, sex and country-adjusted models (Model 1), whereby only smoking (non-smokers and smokers in tertiles of pack-years) proved to be a confounder (>10% change of the risk estimators). The lowest Akaike´s Information Criterion value was reached by further including BMI 2 years before recruitment (normal weight/overweight/obese) (Model 2).
Effect modification by country, center, age, gender, smoking and alcohol status, and BMI variables was evaluated by adding interaction terms in the models, and comparing them with models lacking this interaction (likelihood ratio test, LHR). Effect measure modification was further evaluated in stratified analyses by subgroups of these variables.

Dose-response and trend analysis was conducted by fitting the categorized variables (time since T2DM, age at T2DM diagnosis and Hb1Ac levels) as an ordinal score in the logistic models. The dose-response curve was evaluated by applying restricted cubic splines (3 knots at the 10%, 50% and 90% percentile). Linearity tests were performed by comparing via the LHR test the continuous variable models as nonlinear or as linear. Interaction by centre but not by country was apparent; therefore, random centre effects in mixed models when appropriate were applied.

Mediation analysis: The counterfactual mediation model for binary mediators and outcomes was used to explore mediation effects on the associations. We explored whether obesity leading to T2DM, and subsequently to PC, could explain the observational association between T2DM and PC. With this method, we estimated the total effect (TE) of obesity on PC by determining a natural direct effect (NDE) of obesity on PC and a natural indirect effect (NIE) of obesity on T2DM accounting for the influence of observed confounders. Standard errors were generated using Monte Carlo bootstrapping with 1,000 replications. Similarly, potential mediating effects of obesity on the association between T2DM and PC risk were explored.

*Mendelian Randomizatio*n *Analysis (MRA):* The causal effect of T2DM subtypes on PC (Supplementary Figure 1 B) was estimated using several MRA approaches (Wald ratio, 2-stage least squares -TSLS, inverse variance weighted method-IVW, and simple median), adjusting estimates for the aforementioned potential confounders. Some of these methods were applied via the *MendelianRandomization* R package. A total of 16 variants in high LD (R^2>0.8) were removed for these analyses (Supplementary Table 1). Since genetic variants for T2DM can be confounded by BMI effects due to sharing of variants (i.e., pleiotropy), we tested for the association between the variants and BMI, as well as other confounders, and removed those variants showing an association with other traits (Supplementary Table 3). After removing them, 35 T2DM-SNPs remained to build the IV. The genetic association of this IV with T2DM was estimated in controls only, and subsequently the association with PC was estimated in the full case-control sample. Logistic regression models adjusted for age, sex and five principal components to control for population stratification were used to assess the per allele effect of each SNP and of the genetic score. In addition, the weighted median estimation and the MR-Egger approach were applied to detect and correct bias due to pleiotropy. The weighted median estimator reflects the median of the distribution of weighted Wald ratio estimates. This test is less sensitive to the influence of pleiotropic variants since less weight is given to outlying estimates. The MR-Egger approach performs a weighted linear regression of the genetic associations with the outcome on the genetic associations with the exposure, while keeping the intercept unconstrained. This test provides evidence for directional pleiotropy when the intercept differs from zero.

Bidirectional MRA: The same procedure was used to explore the causal effect of PC on T2DM (Supplementary Figure 1 C). We kept 33 PC-related SNPs for the analyses after removing SNPs in LD and those associated with other traits (Supplementary Tables 1 and 3). The association

of the IV with PC was estimated in individuals without T2DM, followed by its association with T2DM in all subjects.

MRA using pleiotropic genetic variants: Causal assessment of obesity (at age 50 and 2 years before the interview) and PC was explored considering 85 obesity-related SNPs (41 SNPs were removed due to LD and associations with other traits: Supplementary Tables 1 and 3). Multivariable MRA was used to disentangle further the causal effect of T2DM and obesity on PC using T2DM-SNPs as IV (Supplementary Figure 1 D), or PC-SNPs as IV in the opposite direction (Supplementary Figure 1 E). The IVW, TSLS and Egger methods were applied in these analyses. In line with the aforementioned mediation analyses, we explored potential mediating effects of obesity or T2DM (mediators) using separate IVs (Supplementary Figure 1 F and G). Direct and indirect effects were estimated using the counterfactual method.

Sensitivity analyses: We compared estimates from the unimputed and imputed data to assess the robustness of the results. Although heterogeneity by country was absent, we evaluated the consistency of the results across countries by removing each country at a time from the analyses. This was particularly relevant for PC cases from Italy due to the lack of matched controls. Sensitivity analyses also comprised the assessment of T2DM status based on questionnaire, i.e. self-reported (SR) data, or biomarker data in different study settings. In MRA, to further detect potential pleiotropic variants, we also removed SNPs that were outliers based on Cook´s distances and removed additional variants potentially associated with other phenotypes. The latter were identified in publicly available data from GWAS studies (PhenoScanner database).The MR-base platform was also used to inspect the presence of pleiotropy. For instance, scatter plots of the gene-outcome and gene-exposure associations and for the SNP risk increase against the strength of instrumental SNPs were constructed, along with leave-one out analyses funnel plots for visual assessment of pleitropy. In addition, since unmeasured confounding is a major concern of causal inference in observational studies, we tested by estimating the E-Value how strong such confounders would have to be related to the exposure and the outcome to explain away the observed association. High E-Values reflect less impact of these confounders on the observed associations.

Results were comparable to those seen in analyses of the original data, regarding the use of unimputed missing data (Supplementary Tables 14), country-specific data (data not shown), reclassified T2DM status with biomarker data (e.g., Supplementary Table 15), and in analyses of the influence of pleiotropic effects in MRA (Supplementary Table 16 and Supplementary Figure 3). The E-value for the causal effect between NODM or LSDM with PC risk (E-value = 12.29 and 3.12, respectively) suggested that unmeasured confounders are unlikely to explain away the effect of the observed association, especially with regard to NODM (Supplementary Table 17).

Supplementary Tables

Supplemental Table 1. Selected Genetic variants of T2DM, PC, and obesity.

	T2DM			PC			Obesity		
ID	SNP	Chr	Position	SNP	Chr	Position	SNP	Chr	Position
1	rs2641348	1	120437884	rs13303010	1	894573	rs11208659	1	65979280
2	rs340874	1	214159256	rs1747924	1	64538961	rs3101336	1	72751185
3	rs13414140	2	43671176	rs351365	1	113046395	rs2568958	1	72765116
4	rs243021	2	60584819	rs10919791	1	199965168	rs7531118	1	72837239
5	rs2943641	2	227093745	rs2816938	1	199985368	rs1993709	1	72838529
6	rs780094	2	27741237	rs3790844	1	200007432	rs1514177	1	74991402
7	rs1801282	3	12393125	rs962856	2	67593803	rs1514174	1	74993063
8	rs1470579	3	185529080	rs1486134	2	67639769	rs17381664	1	78048331
9	rs4402960	3	185511687	rs12478462	2	153654720	rs12408810	1	106640943
10	rs11708067	3	123065778	rs9854771	3	189508471	rs17024258	1	110147321
11	rs2877716	3	123094451	rs6537481	4	148396094	rs633715	1	177852580
12	rs4411878	3	64703665	rs2736098	5	1294086	rs12130212	1	209727257
13	rs6802898	3	12391207	rs35226131	5	1295373	rs2605100	1	219644224
14	rs10012946	4	6293350	rs401681	5	1322087	rs6429082	1	235600129
15	rs7708285	5	76425867	rs31490	5	1344458	rs12145833	1	243483754
16	rs9472138	6	43811762	rs17688601	7	40866663	rs6711012	2	624034
17	rs1535500	6	39284050	rs73328514	7	47488569	rs12463617	2	629244
18	rs4712523	6	20657564	rs6971499	7	130680521	rs11127485	2	632028
19	rs10946398	6	20661034	rs2941471	8	76470404	rs10189761	2	646364
20	rs7754840	6	20661250	rs10094872	8	128719884	rs10182181	2	25150296
21	rs7766070	6	20686573	rs1561927	8	129568078	rs17025867	2	40578559
22	rs7756992	6	20679709	rs10991043	9	106797388	rs6726292	2	55156630
23	rs13234407	7	130438214	rs2417487	9	106887581	rs6731302	2	58833493
24	rs1635852	7	28189411	rs687289	9	136137106	rs887912	2	59302877
25	rs2191348	7	15064255	chr9_136149229	9	136149229	rs7581710	2	121195181
26	rs4607517	7	44235668	rs7310409	12	121424861	rs16867321	2	181362379
27	rs13266634	8	118184783	chr12_121454622	12	121454622	rs7603514	2	206836612
28	rs3802177	8	118185025	rs9554197	13	28476978	rs2943650	2	227105921
29	rs896854	8	95960511	rs9581943	13	28493997	rs11680012	2	238672425
30	rs2383208	9	22132076	rs9543325	13	73916628	rs12635698	3	16408489
31	rs10811661	9	22134094	chr16_75263661	16	75263661	rs1435703	3	25560231
32	rs10512085	9	81924713	rs7200646	16	86335351	rs13078807	3	85884150
33	rs7903146	10	114758349	rs4795218	17	36078510	rs7638110	3	138903985
34	rs5015480	10	94465559	rs77038344	17	38644214	rs1516725	3	185824004
35	rs1111875	10	94462882	chr17_70400166	17	70400166	rs9816226	3	185834499
36	rs7901695	10	114754088	rs7214041	17	70401476	rs13130484	4	45175691
37	rs11257655	10	12307894	rs1517037	18	56878274	rs10938397	4	45182527
38	rs11603334	11	72432985	rs6073450	20	43086648	rs4833407	4	113311790
39	rs1552224	11	72433098	rs450960	22	18316304	rs10433903	4	118093137
40	rs5215	11	17408630	rs16986825	22	29300306	rs4864201	4	130731284
41	rs5219	11	17409572				rs925642	4	187678866

42	rs10830963	11	92708710
43	rs2237892	11	2839751
44	rs1387153	11	92673828
45	rs2334499	11	1696849
46	rs231362	11	2691471
47	rs1353362	12	71613276
48	rs2612067	12	66170163
49	rs7965349	12	121471931
50	rs1359790	13	80717156
51	rs2028299	15	90374257
52	rs7172432	15	62396389
53	rs4778582	15	80420966
54	rs8042680	15	91521337
55	rs8050136	16	53816275
56	rs9939609	16	53820527
57	rs4430796	17	36098040
58			
59			
60			
61			
62			
63			
64			
65			
66			
67			
68			
69			
70			
71			
72			
73			
74			
75			
76			
77			
78			
79			
80			
81			
82			
83			
84			
85			
86			
87			

rs2307111	5	75003678
rs2112347	5	75015242
rs374748	5	127699375
rs9328321	6	5600438
rs4712652	6	22078615
rs999943	6	33624733
rs2274459	6	33762242
rs2206277	6	50798526
rs987237	6	50803050
rs734597	6	50836279
rs2207139	6	50845490
rs2807278	6	131809920
rs10953454	7	104503813
rs545854	8	10002570
rs17150703	8	9745798
rs17126232	8	17977650
rs4735692	8	76615663
rs10968576	9	28414339
rs1412239	9	28425515
rs16933812	9	36969205
rs2275848	9	95887320
rs10508503	10	16299951
rs16923476	10	23858211
rs7474896	10	37982097
rs10999409	10	72332440
rs2116830	10	78646536
rs11042023	11	8662516
rs297325	11	16389594
rs4756846	11	16403511
rs12295638	11	26605331
rs988712	11	27563382
rs2030323	11	27728539
rs564343	11	65895166
rs1048466	12	551550
rs3782724	12	6466081
rs10875976	12	50226467
rs7138803	12	50247468
rs11109072	12	97901270
rs9568856	13	54064981
rs9568867	13	54107352
rs17081231	13	66967622
rs534870	13	80959207
rs7989336	13	97017548
rs1957894	14	61908111
rs699363	14	72692493
rs11624704	14	78786077

88	rs7141420	14	79899454
89	rs2370983	14	79903376
90	rs8028313	15	68043057
91	rs970843	15	98876029
92	rs2531995	16	4013467
93	rs12446554	16	19935073
94	rs12446632	16	19935389
95	rs11639988	16	19944363
96	rs7498665	16	28883241
97	rs7184597	16	28921809
98	rs1421085	16	53800954
99	rs1558902	16	53803574
100	rs1121980	16	53809247
101	rs17817449	16	53813367
102	rs8043757	16	53813450
103	rs8050136	16	53816275
104	rs7185735	16	53822651
105	rs9941349	16	53825488
106	rs9923451	16	78952439
107	rs1424233	16	79682751
108	rs7187365	16	86511915
109	rs9299	17	46669430
110	rs7503807	17	78591111
111	rs1805081	18	21140432
112	rs17697518	18	38765659
113	rs1631486	18	53026357
114	rs17700144	18	57811982
115	rs538656	18	57850422
116	rs17782313	18	57851097
117	rs10871777	18	57851763
118	rs476828	18	57852587
119	rs11152213	18	57852948
120	rs17773430	18	57963117
121	rs1800437	19	46181392
122	rs10423928	19	46182304
123	rs6110577	20	15335754
124	rs13041126	20	51092996
125	rs11088859	21	22689344
126	rs5762430	22	28378472

SNPs in LD (>0.8) for T2DM-related variants: rs4712523, rs10946398, rs7754840, rs7756992, rs9939609, rs8050136, rs5215, rs1552224, rs11603334, rs5015480, rs7903146, rs10811661, rs3802177, rs1801282, rs2877716, rs1470579

SNPs in LD (>0.8) for PC-related variants: rs31490, rs9554197

SNPs in LD (>0.8) for Obesity-related variants: rs2568958, rs7531118, rs1514174, rs12463617, rs11127485, rs10189761, rs10938397, rs2112347, rs987237, rs734597, rs2207139, rs1412239, rs7184597, rs1558902, rs1121980, rs17817449, rs8043757, rs8050136, rs7185735, rs9941349, rs9923451, rs12446554, rs12446632, rs11639988, rs538656, rs17782313, rs10871777, rs476828, rs11152213, rs10423928

Supplemental Table 2. Evaluation of the performance of the missing data imputation

Variable	Proportion of missings (%)	OOB Error Test[1] in test set	Imputed values	Proportion of concordance ()	OOB Error Test[1] in full set
status	0.00	0.0000	0	**NA**	0.0000
country	0.00	0.0000	0	**NA**	0.0000
gender	0.11	0.1315	2	**50**	0.1320
smoking status	3.29	0.0103	30	**96.67**	0.0015
alcohol status	1.83	0.4893	21	**57.14**	0.4117
chronic pancreatitis status	3.71	0.0052	37	**100**	0.0053
diabetes by type	2.67	0.0031	26	**100**	0.0003
educational level	2.22	0.6254	21	**47.62**	0.5677
FH of pancreatic cancer	3.57	0.0576	26	**100**	0.0498
FH of diabetes[2]	20.15	0.4210	226	**100**	0.3537
periodontitis	26.39	0.1903	257	**81.71**	0.1745
recession	37.16	0.4041	362	**67.96**	0.2862
diabetes diet control	3.99	0.0031	41	**100**	0.0000
diabetes oral medication	3.23	0.0021	42	**97.62**	0.0000
diabetes insulin control	3.65	0.0021	29	**100**	0.0000
Pancreatitis type	3.71	0.0294	46	**93.48**	0.0330
asthma status	9.33	0.0215	113	**100**	0.0015
nasal allergies	8.63	0.0088	94	**98.94**	0.0018
cancer	8.12	0.1393	86	**90.70**	0.1367
diabetes status	1.60	0.0000	15	**100.00**	0.0000
metabolic syndrome	18.27	0.0061	182	**97.25**	0.0007
center	0.00	0.0000	0	**NA**	0.0000
weighton body site	8.68	0.0621	96	**91.67**	0.0160
BMI 2 years before	4.86	0.0000	42	**100**	0.0000
BMI at age 20	19.93	0.0447	192	**95.31**	0.0074
BMI at age 50	27.18	0.0351	257	**94.55**	0.0108
pack-years in tertiles	10.20	0.0055	90	**96.67**	0.0006
age in categories	0.37	0.0643	2	**100**	0.0220
place fat deposition	8.68	0.0132	92	**97.83**	0.0031
weight gain 5 kg	32.66	0.0029	309	**95.15**	0.0008
weight gain 10 kg	32.66	0.0000	324	**94.75**	0.0004
weight at age 20 in tertiles	18.94	0.0109	173	**98.27**	0.0010
weight at age 50 in tertiles	25.74	0.0202	257	**94.55**	0.0019
Weight since age 50	26.45	0.0000	250	**98.00**	0.0000
hypertension	9.19	0.0044	96	**94.79**	0.0009
cholesterol	10.85	0.0000	116	**98.28**	0.0022
height in tertiles	2.22	0.0206	27	**100**	0.0040
smoking duration in tertiles	9.47	0.0143	90	**97.78**	0.0053
smoking intensity in tertiles	3.20	0.0145	31	**100**	0.0044

NA=not applicable; FH=family history

Covariates used to improve imputation were case-control status, country, center, medical history (cancer, asthma, allergies, chronic pancreatitis), smoking variables (intensity and duration) and weight and height.

[1] Out of bag (OOB) error rates: normalized squared error for continuous variables (e.g. age) and proportion of falsely classified entries for categorical variables. Values close to zero indicated good performance and values close to one indicated bad performance.

[2] Concordance test applied to the study population without Ireland since this country did not collect information on family history of the disease.

Supplemental Table 3. Genetic variants associated individually at p-value level <0.05 with T2DM and PC, as well as with selected covariates in the study population (752 controls).

T2DM-SNPs				
NODM	**LSDM**	**PC**	**smoking**	**obesity**
any	rs2943641 (p=0.018) rs1801282 (p=0.051) rs7901695 (p=0.045) rs7903146 (p=0.016)	rs2191348 (p=0.015) rs13266634 (p=0.006) rs3802177 (p=0.005) rs7965349 (p=0.011)	rs2641348 (p=0.006) rs13234407 (p=0.011) rs1111875 (p=0.018) rs5015480 (p=0.020) rs2334499 (p=0.043)	rs10830963 (p=0.044) rs4430796 (p=0.031)
PC-SNPs				
NODM	**LSDM**	**PC**	**smoking**	**obesity**
rs2816938.199985368.T.A (p=0.006) rs7310409 (p=0.016) chr12_121454622_C_T (p=0.005)	any	rs351365.113046395 (p=0.002) rs2816938.199985368 (p=0.001) rs1486134.67639769 (p=0.049) rs31490 (p=0.029) rs73328514.47488569 (p=0.024) rs6971499 (p=0.007) rs2941471.76470404 (p=0.031) rs9543325 (p=0.003) chr16_75263661 (p=0.001) chr17_70400166 (p=0.005) rs7214041.70401476 (p=0.005)	rs6537481.148396094 (p=0.001)	rs1747924.64538961 (p=0.011) rs2816938.199985368 (p=0.012) rs2736098.1294086 (p=0.005) rs17688601.40866663 (p=0.026)

NA: not applicable

Supplemental Table 4. General characteristics of the study population. PanGenEU study (2,018 cases and 1,540 controls). Imputed data

		Cases N=2018		Controls N=1540				
		N	%	N	%	p-value	OR	[95%CI]
Country						<0.001		
	Spain	884	43.80	770	50.00			
	England	126	6.24	22	1.43			
	Germany	131	6.49	111	7.21			
	Ireland	173	8.57	290	18.80			
	Italy	533	26.40	0	0.00			
	Sweden	171	8.47	347	22.50			
Gender						0.143		
	females	873	43.30	705	45.80		Ref.	
	males	1145	56.70	835	54.20		1.11	[0.97;1.27]
Age (years)		64.3	12.10	66.8	12.50	<0.001	0.98	[0.98;0.99]
Age in categories						<0.001		
	<55 y	409	20.30	261	16.90		Ref.	
	55-65 y	500	24.80	323	21.00		0.99	[0.80;1.22]
	65-75 y	708	35.10	500	32.50		0.90	[0.74;1.10]
	≥75 y	401	19.90	456	29.60		0.56	[0.46;0.69]
BMI 2 years before						0.971		
	<25	761	37.70	575	37.30		Ref.	
	25-29.99	868	43.00	668	43.40		0.98	[0.85;1.14]
	≥30	389	19.30	297	19.30		0.99	[0.82;1.19]
BMI at age 20						0.263		
	<25	1750	86.70	1330	86.40		Ref.	
	25-29.99	228	11.30	189	12.30		0.92	[0.75;1.13]
	≥30	40	1.98	21	1.36		1.44	[0.85;2.51]
BMI at age 50						<0.001		
	<25	620	30.70	612	39.70		Ref.	
	25-29.99	929	46.10	604	39.20		1.52	[1.30;1.77]
	≥30	468	23.20	324	21.00		1.43	[1.19;1.71]
Weight gain >5kg (age 20-50)						0.343		
	no	391	19.40	319	20.70		Ref.	
	yes	1627	80.60	1221	79.30		1.09	[0.92;1.28]
Weight gain >10kg (age 20-50)						<0.001		
	no	742	36.80	676	43.90		Ref.	
	yes	1276	63.20	864	56.10		1.35	[1.18;1.54]
Weight loss since age 50						0.012		
	no	1346	66.70	1089	70.70		Ref.	
	yes	672	33.30	451	29.30		1.21	[1.04;1.39]
Smoking status						<0.001		
	never	760	37.70	691	44.90		Ref.	
	former	683	33.80	563	36.60		1.10	[0.95;1.28]
	current	575	28.50	286	18.60		1.83	[1.53;2.18]
Pack-years in tertiles						<0.001		
	never smokers	760	37.70	691	44.90		Ref.	
	[0.05,12.95]	259	12.80	269	17.50		0.88	[0.72;1.07]
	[13,36]	583	28.90	327	21.20		1.62	[1.37;1.92]
	[36.3,240]	416	20.60	253	16.40		1.49	[1.24;1.80]
Alcohol status						<0.001		
	never	599	29.70	390	25.30		Ref.	

	former	508	25.20	234	15.20		1.41	[1.16;1.73]
	current	911	45.10	916	59.50		0.65	[0.55;0.76]
Pancreatitis type						<0.001		
	no	1918	95.00	1523	98.90		Ref.	
	acute	81	4.01	15	0.97		4.25	[2.51;7.71]
	chronic	19	0.94	2	0.13		7.06	[2.03;48.0]
Educational level (years of education)						<0.001		
	<5 y	343	17.00	177	11.50		Ref.	
	6 to 9 y	486	24.10	399	25.90		0.63	[0.50;0.79]
	10 to 13 y	698	34.60	508	33.00		0.71	[0.57;0.88]
	≥14 y	491	24.30	456	29.60		0.56	[0.44;0.69]
Family history of PC						<0.001		
	no	1889	93.60	1499	97.30		Ref.	
	yes	129	6.39	41	2.66		2.49	[1.76;3.60]
Periodontitis						0.643		
	no	1744	86.40	1340	87.00		Ref.	
	yes	274	13.60	200	13.00		1.05	[0.87;1.28]
Recession						0.003		
	no	1481	73.40	1197	77.70		Ref.	
	yes	537	26.60	343	22.30		1.27	[1.08;1.48]
Asthma						<0.001		
	no	1887	93.50	1381	89.70		Ref.	
	yes	131	6.49	159	10.30		0.60	[0.47;0.77]
Nasal allergies						<0.001		
	no	1771	87.80	1236	80.30		Ref.	
	yes	247	12.20	304	19.70		0.57	[0.47;0.68]
Hypertension						<0.001		
	no	1324	65.60	913	59.30		Ref.	
	yes	694	34.40	627	40.70		0.76	[0.67;0.88]
Cholesterol						<0.001		
	no	1459	72.30	1000	64.90		Ref.	
	yes	559	27.70	540	35.10		0.71	[0.61;0.82]

Differences between cases and controls evaluated via Chi-squared test (categprocal variables) and Student´s t-test or Mann-Whitney (continuous variables).
Odds Ratios (OR) derived from unadjusted unconditional logistic resgression models.
Data of all variables was self-reported.

Supplemental Table 5. Baseline characteristics of NODM and LSDM in the PanGenEU study (538 cases and 198 controls). Imputed data.

	LSD M N=509 N	%	NODM N=227 N	%	*p-value*	OR NODM *vs* LSDM	95%CI
Age (years)					<0.001		
<55	37	7.27	28	12.30		Ref.	
55-65	92	18.10	67	29.50		1.04	[0.58;1.86]
65-75	213	41.80	83	36.60		1.94	[1.11;3.37]
≥75	167	32.80	49	21.60		2.57	[1.42;4.63]
Gender					0.818		
females	180	35.40	83	36.60		Ref.	
males	329	64.60	144	63.40		1.05	[0.76;1.46]
Smoking status					0.441		
never	199	39.10	78	34.40		Ref.	
former	201	39.50	94	41.40		0.84	[0.58;1.20]
current	109	21.40	55	24.20		0.78	[0.51;1.18]
Alcohol status					0.127		
never	155	30.50	57	25.10		Ref.	
former	140	27.50	78	34.40		0.66	[0.44;1.00]
current	214	42.00	92	40.50		0.86	[0.58;1.26]
Chronic pancreatitis					0.298		
no	504	99.00	222	97.80		Ref.	
yes	5	0.98	5	2.20		0.44	[0.12;1.65]
Educational level (years)					0.632		
<5 y	114	22.40	44	19.40		Ref.	
6 to 9 y	138	27.10	66	29.10		0.81	[0.51;1.27]
10 to 13 y	164	32.20	69	30.40		0.92	[0.58;1.43]
≥14 y	93	18.30	48	21.10		0.75	[0.46;1.23]
Family history PC					0.318		
no	478	93.90	218	96.00		Ref.	
yes	31	6.09	9	3.96		1.55	[0.75;3.54]
Family history Diabetes					0.667		
no	252	49.50	117	51.50		Ref.	
yes	257	50.50	110	48.50		1.08	[0.79;1.48]
Periodontitis					0.221		
no	421	82.70	197	86.80		Ref.	
yes	88	17.30	30	13.20		1.37	[0.88;2.17]
Recession					0.667		
no	381	74.90	174	76.70		Ref.	
yes	128	25.10	53	23.30		1.1	[0.77;1.60]
Diabetes age diagnosis					<0.001		
≤ 55y	207	40.70	37	16.30		Ref.	
55 to ≤ 65y	167	32.80	63	27.80		0.48	[0.30;0.75]
> 65y	135	26.50	127	55.90		0.19	[0.12;0.29]
Diabetes with diet					0.770		
yes	395	77.60	174	76.70		Ref.	
no	114	22.40	53	23.30		0.95	[0.65;1.38]
Diabetes with oral medication					<0.001		
yes	407	80.00	139	61.20		Ref.	
no	102	20.00	88	38.80		0.4	[0.28;0.56]
					0.916		

Diabetes with insulin							
	yes	293	57.60	129	56.80	Ref.	
	no	216	42.40	98	43.20	0.97	[0.71;1.33]
Asthma					0.135		
	no	467	91.70	216	95.20	Ref.	
	yes	42	8.25	11	4.85	1.75	[0.91;3.65]
Nasal allergies					0.46		
	no	453	89.00	197	86.80	Ref.	
	yes	56	11.00	30	13.20	0.81	[0.51;1.32]
Metabolic syndrome					0.116		
	Any one	146	28.70	58	25.60	Ref.	
	Any two	170	33.40	96	42.30	0.7	[0.47;1.04]
	Any three	151	29.70	60	26.40	1	[0.65;1.53]
	All four	42	8.25	13	5.73	1.27	[0.65;2.64]
BMI 2 years before					0.172		
	<25	128	25.10	47	20.70	Ref.	
	25-29.99	245	48.10	126	55.50	0.72	[0.48;1.06]
	≥30	136	26.70	54	23.80	0.93	[0.58;1.47]
BMI at age 20					0.713		
	<25	421	82.70	184	81.10	Ref.	
	25-29.99	72	14.10	37	16.30	0.85	[0.55;1.32]
	≥30	16	3.14	6	2.64	1.15	[0.46;3.29]
BMI at age 50					0.99		
	<25	110	21.60	48	21.10	Ref.	
	25-29.99	241	47.30	108	47.60	0.97	[0.64;1.46]
	≥30	158	31.00	71	31.30	0.97	[0.62;1.51]
Pack-years in tertiles					0.585		
	never smokers	199	39.10	78	34.40	Ref.	
	[0.05,12.95]	51	10.00	27	11.90	0.74	[0.43;1.28]
	[13,36]	129	25.30	64	28.20	0.79	[0.53;1.18]
	[36.3,240]	130	25.50	58	25.60	0.88	[0.59;1.32]
Weightgain >5kg (age 20-50)					0.943		
	no	58	11.40	27	11.90	Ref.	
	yes	451	88.60	200	88.10	1.05	[0.64;1.70]
Weightgain >10kg (age 20-50)					0.797		
	no	155	30.50	72	31.70	Ref.	
	yes	354	69.50	155	68.30	1.06	[0.75;1.48]
Hypertension					0.596		
	no	241	47.30	113	49.80	Ref.	
	yes	268	52.70	114	50.20	1.1	[0.81;1.51]
Cholesterol					0.981		
	no	316	62.10	140	61.70	Ref.	
	yes	193	37.90	87	38.30	0.98	[0.71;1.36]
Weight loss since age 50					0.429		
	no	333	65.4	156	68.7	Ref.	
	yes	176	34.6	71	31.3	1.16	[0.83;1.63]

Differences between cases and controls evaluated via Chi-squared test (categprocal variables) and Student´s t-test (continuous variables).
Odds Ratios (OR) derived from unadjusted unconditional logistic resgression models.

Supplemental Table 6. Association between diabetes-related variables and PC risk in the PanGenEU study (2,018 cases and 1,540 controls) when adjusting for T2DM treatment and duration of the disease.

	Age, sex, country-adjusted (Model 1)		Model 1 + use of oral medication		Model 1 + use of insulin		Model 1 + duration of diabetes	
	OR	**[95%CI]**	**OR**	**[95%CI]**	**OR**	**[95%CI]**	**OR**	**[95%CI]**
Diabetes status:								
no diabetes	Ref.		Ref.		Ref.		Ref.	
yes	2.56	[2.10;3.11]	3.01	[2.13;4.26]	1.77	[1.37;2.29]	1.19	[0.76;1.85]
Diabetes status by subtype								
no diabetes	Ref.		Ref.		Ref.		Ref.	
yes, ≤ 2 years (NODM)	6.49	[4.25;9.90]	6.36	[3.95;10.26]	4.51	[2.88;7.06]	2.64	[1.4;4.97]
yes, > 2years (LSDM)	1.90	[1.53;2.37]	1.85	[1.25;2.73]	1.27	[0.96;1.69]	1.19	[0.76;1.85]
Family history of diabetes[1]								
no diabetes	Ref.		Ref.		Ref.		Ref.	
yes	1.25	[1.05;1.49]	1.07	[0.89;1.28]	1.07	[0.89;1.28]	1.07	[0.89;1.28]
Diabetes by age at diagnosis (years)[2]								
no diabetes	Ref.		Ref.		Ref.		Ref.	
≤ 55 y	1.56	[1.15;2.11]	1.83	[1.21;2.76]	0.91	[0.62;1.32]	1.17	[0.75;1.83]
55 to ≤ 65 y	2.68	[1.92;3.73]	3.23	[2.02;5.15]	1.81	[1.25;2.61]	1.50	[0.79;2.86]
> 65 y	4.06	[2.94;5.60]	4.75	[3.13;7.28]	2.79	[1.96;3.97]		
	p-trend	2E-16	*p-trend*	1.2E-05	*p-trend*	6.6E-07	*p-trend*	0.122
Diabetes by time since diagnosis (years)[2]								
no diabetes	Ref.		Ref.		Ref.		Ref.	
≤1 y	11.14	[6.09;20.37]	11.04	[5.83;20.93]	7.24	[3.89;13.5]	NA	NA
1 to ≤2 y	2.64	[1.40;4.97]	2.61	[1.3;5.22]	1.72	[0.89;3.35]	NA	NA
2 to ≤5 y	2.40	[1.54;3.73]	2.36	[1.35;4.14]	1.73	[1.09;2.76]	NA	NA

5 to ≤10 y	2.67	[1.81;3.93]	2.63	[1.57;4.47]	1.73	[1.14;2.66]	NA	NA
10 to ≤20 y	1.59	[1.09;2.32]	1.56	[0.94;2.62]	0.91	[0.58;1.42]	NA	NA
>20 y	1.19	[0.76;1.85]	1.17	[0.69;1.98]	0.61	[0.36;1.04]	NA	NA
	p-trend	9.6E-08	*p-trend*	5.5E-10	*p-trend*	1.4E-11	*p-trend*	NA
Diabetes controlled with diet								
no diabetes	Ref.		Ref.		Ref.		Ref.	
yes	2.60	[2.09;3.24]	3.08	[2.14;4.43]	1.75	[1.32;2.32]	1.22	[0.78;1.92]
no use	2.43	[1.69;3.49]	2.84	[1.81;4.47]	1.82	[1.24;2.69]	1.04	[0.6;1.82]
Use of oral medication								
no diabetes	Ref.		Ref.		Ref.		Ref.	
yes	2.41	[1.93;3.00]	NA	NA	1.74	[1.33;2.28]	1.19	[0.75;1.9]
no use	3.01	[2.13;4.26]	NA	NA	1.89	[1.25;2.85]	1.17	[0.69;1.98]
Use of insulin								
no diabetes	Ref.		Ref.		Ref.		Ref.	
yes	3.76	[2.85;4.95]	3.96	[2.71;5.79]	NA	NA	1.61	[1.01;2.58]
no use	1.77	[1.37;2.29]	1.89	[1.25;2.85]	NA	NA	0.61	[0.36;1.04]

[1] Information on family history of diabetes was not collected in Ireland; results are based on data for 1,845 cases and 1,250 controls

[2] Linear associations for age since T2DM diagnosis and nonlinear association for time since T2DM (Supplemental Figure 1)

Model 1: adjusted for age (<55, 55-65, 65-75, ≥75 years), gender, country.

Model 2: Model 1 also adjusted for use of oral medication.

Model 3: Model 1 also adjusted for use of insulin.

Model 4: Model 1 also adjusted for duration of T2DM.

NA=not applicable

Supplemental Table 7. Association between T2DM status based on Hb1Ac levels and questionnaire data and PC risk in the PanGenEU study.

	Cases		Controls			Crude Model		Model1		Model2	
	P50	IQR	P50	IQR	p-value[1]	OR	[95%CI]	OR	[95%CI]	OR	[95%CI]
HbA1c (%)[2,3] per 1 unit increase	6.1	5.6;6.9	5.6	5.4;6.0	<0.001	1.48	[1.30;1.69]	1.50	[1.31;1.71]	1.49	[1.30;1.70]
C-Peptide[2,3] per log2 increase	2.3	1.4;3.7	4.2	2.5;6.4	<0.001	0.45	[0.38;0.56]	0.46	[0.39;0.54]	0.46	[0.39;0.53]
	N	%	N	%							
Diabetogenic status by HbA1c levels[3]					<0.001						
HbA1c <6.5%	336	66	354	85.7		Ref.		Ref.		Ref.	
HbA1c ≥6.5%	173	34	59	14.3		3.08	[2.22;4.32]	3.29	[2.34;4.62]	3.27	[2.32;4.60]
Biomarker and self-reported diabetes status					<0.001						
no diabetes	286	56.2	322	78		Ref.		Ref.		Ref.	
self-reported but normal Hb1Ac levels	50	9.8	32	7.7		1.76	[1.10;2.84]	1.9	[1.18;3.10]	1.92	[1.19;3.13]
self-reported and HbA1c ≥6.5%	173	34	59	14.3		3.3	[2.37;4.65]	3.59	[2.55;5.11]	3.58	[2.53;5.11]
Diabetes status[3]					<0.001						
no diabetes	350	68.7	341	82.6		Ref.		Ref.		Ref.	
yes	159	31.3	72	17.40		2.15	[1.57;2.96]	2.26	[1.64;3.15]	2.26	[1.64;3.15]
Reclassified diabetes status[3]					<0.001						
no diabetes	286	56.2	322	78		Ref.		Ref.		Ref.	
self-reported and/or HbA1c ≥6.5%	223	43.8	91	22		2.76	[2.07;3.70]	2.99	[2.21;4.06]	2.99	[2.21;4.07]
Diabetes status by subtype[3]					<0.001						
no diabetes	350	68.7	341	82.6		Ref.		Ref.		Ref.	
NODM	66	13	15	3.6		4.29	[2.47;7.93]	4.53	[2.59;8.42]	4.58	[2.61;8.55]
LSDM	93	18.3	57	13.8		1.59	[1.11;2.29]	1.65	[1.14;2.41]	1.64	[1.13;2.40]
Reclassified diabetes status by subtypes[3]					<0.001						
no diabetes	286	56.2	322	78		Ref.		Ref.		Ref.	
NODM	130	25.5	34	8.2		4.3	[2.89;6.57]	4.63	[3.08;7.12]	4.63	[3.07;7.15]
LSDM	93	18.3	57	13.8		1.84	[1.28;2.66]	1.98	[1.35;2.90]	1.97	[1.35;2.90]
Biomarker Hb1Ac levels[3]					<0.001						
<5.5	100	19.6	129	31.2		Ref.		Ref.		Ref.	

5.5-5.8	72	14.1	121	29.3		0.77	[0.52;1.14]		0.71	[0.47;1.06]		0.71	[0.47;1.06]
5.8-6.0	50	9.9	51	12.5		1.26	[0.79;2.03]		1.26	[0.78;2.04]		1.23	[0.76;1.99]
6.0-6.5	114	22.4	53	13		2.76	[1.83;4.22]		2.75	[1.80;4.24]		2.72	[1.77;4.17]
≥6.5	173	34	59	14		3.77	[2.55;5.62]		4.03	[2.69;6.08]		3.99	[2.64;6.01]
					p-trend	2.70E-10		*p-trend*	2.00E-16		*p-trend*	2.00E-16	
Reclassified NODM into type 3c-like diabetes[3]													
no diabetes	286	56.20	322	77.97		Ref.			Ref.			Ref.	
NODM and C-Peptide >4.2 µg/L	37	7.20	21	5.08		1.98	]1.15;3.52]		2.30	[1.31;4.13]		2.28	[1.30;4.10]
NODM and C-Peptide <4.2 µg/L (T3c)	93	18.30	13	3.15		8.05	[4.57;15.37]		8.31	[4.69;15.93]		8.38	[4.71;16.11]
LSDM	93	18.30	57	13.80		1.84	[1.28;2.66]		1.99	[1.36;2.92]		1.98	[1.35;2.92]
Diabetes status[4]					<0.001								
no diabetes	1480	73.3	1342	87.1		Ref.			Ref.			Ref.	
yes	538	26.7	198	12.9		2.46	[2.06;2.95]		2.56	[2.10;3.11]		2.5	[2.05;3.05]
Reclassified diabetes status[4]					<0.001								
no diabetes	1416	70.2	1323	85.9		Ref.			Ref.			Ref.	
self-reported and/or HbA1c ≥6.5%	602	29.8	217	14.1		2.59	[2.18;3.08]		2.85	[2.36;3.45]		2.79	[2.31;3.39]
Diabetes status by subtype PanGenEU[4]					<0.001								
no diabetes	1480	73.3	1342	87.1		Ref.			Ref.			Ref.	
NODM	200	9.91	27	1.75		6.68	[4.52;10.3]		6.49	[4.25;9.90]		6.39	[4.18;9.78]
LSDM	338	16.7	171	11.1		1.79	[1.47;2.19]		1.9	[1.53;2.37]		1.86	[1.49;2.32]
Reclassified diabetes status by subtypes[4]					<0.001								
no diabetes	1416	70.2	1323	85.9		Ref.			Ref.			Ref.	
NODM	264	13.1	46	3		5.36	[3.93;7.49]		5.74	[4.14;8.11]		5.67	[4.09;8.03]
LSDM	338	16.7	171	11.1		1.85	[1.52;2.26]		2.03	[1.62;2.53]		1.98	[1.59;2.48]
Reclassified NODM into type 3c-like diabetes[4]					<0.001								
no diabetes	1416	70.2	1323	85.9		Ref.			Ref.			Ref.	
NODM and C-Peptide >4.2 µg/L	171	8.5	33	2.2		4.84	[3.36;7.20]		4.65	[3.16;7.02]		4.51	[3.06;6.83]
NODM and C-Peptide <4.2 µg/L (T3c)	93	4.6	13	0.8		6.68	[3.86;12.58]		8.83	[5.06;1.67]		8.86	[5.07;1.68]
LSDM	338	16.7	171	11.1		1.85	[1.52;2.26]		2.06	[1.65;2.56]		2.00	[1.61;2.51]
diabetes status[5]					<0.001								

no diabetes	596	67.4	628	81.6		Ref.		Ref.		Ref.		
yes	288	32.6	142	18.4		2.14	[1.70;2.70]	2.09	[1.64;2.66]	2.07	[1.62;2.64]	
Reclassified diabetes status[5]					<0.001							
no diabetes	532	60.2	609	19.1		Ref.		Ref.		Ref.		
self-reported and/or HbA1c ≥6.5%	352	39.8	161	80.9		2.5	[2.01;3.12]	2.57	[2.04;3.24]	2.54	[2.01;3.22]	
Diabetes status by subtype[5]					<0.001							
no diabetes	596	67.4	628	81.6		Ref.		Ref.		Ref.		
NODM	109	12.2	20	2.6		5.74	[3.60;9.63]	5.7	[3.54;9.62]	5.67	[3.52;9.60]	
LSDM	179	20.3	122	15.8		1.55	[1.20;2.00]	1.47	[1.12;1.93]	1.45	[1.11;1.91]	
Reclassified diabetes status by subtypes[5]					<0.001							
no diabetes	532	60.2	609	79.1		Ref.		Ref.		Ref.		
NODM	173	19.6	39	5.1		5.08	[3.56;7.42]	5.4	[3.75;7.94]	5.35	[3.71;7.88]	
LSDM	179	20.2	122	15.8		1.68	[1.30;2.18]	1.63	[1.24;2.15]	1.62	[1.23;2.14]	
Reclassified NODM into type 3c-like diabetes[5]												
no diabetes	532	60.2	609	79.1		Ref.		Ref.		Ref.		
NODM and C-Peptide >4.2 µg/L	80	9.1	26	3.4		3.52	[2.26;5.66]	3.49	[2.21;5.68]	3.41	[2.15;5.57]	
NODM and C-Peptide <4.2 µg/L (T3c)	93	10.5	13	1.7		8.19	[4.70;15.50]	9.46	[5.40;17.96]	9.47	[5.40;18.02]	
LSDM	179	20.2	122	15.8		1.68	[1.30;2.18]	1.65	[1.26;2.18]	1.63	[1.24;2.16]	

[1] Differences between groups evaluated by the Chi-square test (categorical variables) and Mann-Whitney test (continuous variables).
[2] Linear association for Hb1Ac levels and non-linear for C-Peptide (Supplemental Figure 1).
[3] NODM and LSDM was classified with questionnaire and biomarker data in the biomarker study population (509 cases and 413 controls).
[4] NODM and LSDM was classified with questionnaire and biomarker data in the entire study population (2,018 cases and 1,540 controls).
[5] NODM and LSDM was classified with questionnaire and biomarker data in the PanGenEU-Spain study population (884 cases and 770 controls).
Crude Model: unadjusted.
Model 1: adjusted for age (<55, 55-65, 65-75, ≥75 years), sex and center (Spain) or country.
Model 2: Model 1 also adjusted for pack years (never-smokers and tertiles of pack-years) and BMI (<25, 25-30, ≥30 kg/m2).

Supplemental Table 8. Association between T2DM and PC risk by T2MD subtypes and other covariates in the PanGenEU study (2,018 cases and 1,540 controls).

	No diabetes (Ref.)	NODM		LSDM	
	Cases;Controls	Cases;Controls	OR[1] [95%CI]	Cases;Controls	OR[1] [95%CI]
Alcohol status					
never	451;326	47;10	3.62 [1.81;7.90]	101;54	1.43 [0.95;2.16]
former	329;195	71;7	6.59 [3.09;16.33]	108;32	2.61 [1.66;4.20]
current	700;821	82;10	8.21 [4.30;17.34]	129;85	1.71 [1.24;2.37]
	p-value for interaction	0.3530		0.3914	
Chronic pancreatitis					
no	1469;1342	195;27	6.31 [4.2;9.83]	335;169	1.92 [1.55;2.40]
yes	11;0	5;0	NA	3;2	NA
	p-value for interaction	NA		NA	
Family history PC					
no	1378;1305	191;27	6.26 [4.16;9.77]	311;167	1.88 [1.50;2.35]
yes	93;37	9;0	NA	27;4	2.79 [0.90;10.72]
	p-value for interaction	0.2209		0.4220	
Asthma					
no	1378;1207	192;24	6.83 [4.46;10.9]	317;150	2 [1.59; 2.52]
yes	102;135	8;3	2.56 [0.6;13.26]	21;21	1.34 [0.61; 2.9]
	p-value for interaction	0.1853		0.1832	
Nasal allergies					
no	1292;1065	172;25	5.5 [3.59; 8.75]	307;146	1.84 [1.46; 2.34]
yes	188;277	28;2	17.39 [4.92; 110.72]	31;25	2.1 [1.14; 3.88]
	p-value for interaction	0.1131		0.8289	
BMI 2 years before					
<25	635;524	45;2	20.35 [6.11;126.34]	79;49	1.22 [0.79;1.87]
25-29.99	593;570	109;17	5.41 [3.2;9.65]	165;80	2.16 [1.57;2.99]
≥30	252;248	46;8	5.01 [2.34;12.04]	94;42	2.37 [1.53;3.71]
	p-value for interaction	0.1222		0.0496	
BMI at age 20					
<25	1311;1162	163;21	6.72 [4.26;11.12]	271;147	1.78 [1.40;2.26]
25-29.99	143;161	31;6	5.06 [2.08;14.28]	53;19	3.11 [1.69;5.93]
≥30	26;19	6;0	NA	11;5	2.69 [0.63;13.71]
	p-value for interaction	0.5767		0.3708	
BMI at age 50					
<25	516;562	39;9	4.21 [2.04;9.57]	66;44	1.58 [1.02;2.44]
25-29.99	665;517	100;8	11.03 [5.55;25.21]	164;77	1.93 [1.39;2.70]
≥30	299;263	61;10	4.98 [2.48;10.92]	108;50	2.14 [1.40;3.28]
	p-value for interaction	0.1333		0.635	
Age categorized:					
<55	356;249	24;4	4.73 [1.69;16.79]	29;8	3.17 [1.42;7.78]
55-65	369;295	61;6	7.94 [3.56;21.15]	70;22	2.20 [1.28;3.86]
65-75	482;430	75;8	7.54 [3.72;17.45]	151;62	2.18 [1.54;3.11]

≥75	273;368	40;9	4.74 [2.27;10.89]	88;79	1.34 [0.93;1.94]
	p-value for interaction	0.7985		0.2534	
Place fat deposition					
Never carried any extra weight	90;72	4;1	4.95 [0.59;111.11]	10;4	2.71 [0.77;11.19]
Abdominal	966;927	140;20	6.24 [3.87;10.53]	239;134	1.78 [1.38;2.30]
Hips	127;120	10;2	7.52 [1.66;53.69]	12;5	4.94 [1.55;18.11]
All over equally	297;223	46;4	8.9 [3.43;30.4]	77;28	2.05 [1.24;3.47]
	p-value for interaction	0.8896		0.5682	
Weight gain >5kg between age 20-50					
no	326;299	26;1	19.9 [4.08;359.30]	39;19	1.64 [0.89;3.10]
yes	1154;1043	174;26	6.12 [4.02;9.67]	299;152	2.04 [1.61;2.59]
	p-value for interaction	0.1841		0.4969	
Weight gain >10kg between age 20-50					
no	578;613	62;10	6.08 [3.17;12.90]	102;53	2.06 [1.42;3.01]
yes	902;729	138;17	6.62 [3.98;11.66]	236;118	1.87 [1.42;2.46]
	p-value for interaction	0.8334		0.7456	

[1] Odds ratios (ORs) adjusted for age (<55, 55-65, 65-75, ≥75 years), gender, country, pack years (never-smokers and tertiles of pack-years) and BMI (<25, 25-30, ≥30 kg/m^2), except: sex in analyses stratified by sex, BMI in analyses stratified by obesity and pack-years in analyses stratified by smoking status.
[2] Obesity status defined based on BMI 2 years before recruitment.

Supplemental Table 9. Factors associated with PC risk among patients with NODM and LSDM in the PanGenEU study (2,018 cases and 1,540 controls).

		NODM (N=227)				LSDM (N=509)				NODM SR + BIOM (N=310)			
		Cases; Controls	OR[1]	[95%CI]	*p-value*	Cases; Controls	OR[1]	[95%CI]	*p-value*	Cases; Controls	OR[1]	[95%CI]	*p-value*
Gender	females	67;16	Ref.			115;65	Ref.			103;22	Ref.		
	males	133;11	2.59	[1.07;6.32]	0.04	223;106	1.07	[0.7;1.63]	0.75	161;24	1.2	[0.62;2.35]	0.59
Age groups	<65y	85;10	Ref.			99;30	Ref.			106;17	Ref.		
	≥65y	115;17	3.94	[0.8;19.44]	0.09	239;141	1.33	[0.67;2.67]	0.42	158;29	1.73	[0.54;5.60]	0.36
Pbese[2]	no	154;19	Ref.			244;129	Ref.			207;32	Ref.		
	yes	46;8	0.6	[0.22;1.61]	0.31	94;42	1.22	[0.78;1.93]	0.39	57;14	0.56	[0.27;1.19]	0.13
Smoking status	never	67;11	Ref.			120;79	Ref.			96;18	Ref.		
	former	87;7	1.03	[0.32;3.32]	0.96	139;62	1.46	[0.88;2.40]	0.14	108;15	0.87	[0.36;2.1]	0.76
	current	46;9	0.52	[0.15;1.74]	0.29	79;30	1.25	[0.69;2.28]	0.46	60;14	0.56	[0.22;1.46]	0.24
Alcohol status	never	47;10	Ref.			101;54	Ref.			69;16	Ref.		
	former	71;7	1.97	[0.59;6.60]	0.27	108;32	1.91	[1.06;3.43]	0.03	91;13	1.64	[0.66;4.06]	0.29
	current	82;10	1	[0.29;3.45]	1	129;85	0.76	[0.45;1.3]	0.32	104;17	1.27	[0.51;3.14]	0.61
Family history PC	no	191;27	Ref.			311;167	Ref.			252;44	Ref.		
	yes	9;0	NA		NA 1	27;4	3.98	[1.33;11.93]	0.01	12;2	1.34	[0.27;6.61]	0.72
Family history diabetes	no	105;12	Ref.			161;90	Ref.			154;26	Ref.		
	yes	95;15	1.02	[0.42;2.47]	0.97	176;81	1.34	[0.89;2.02]	0.16	110;20	1.07	[0.54;2.10]	0.84
Periodontitis	no	175;22	Ref.			280;141	Ref.			234;37	Ref.		
	yes	25;5	0.48	[0.15;1.57]	0.22	58;30	1.01	[0.60;1.70]	0.98	30;9	0.5	[0.2;1.22]	0.13
Diabetes with diet[3]	yes	156;18	Ref.			264;131	Ref.			156;18	Ref.		
	no	44;9	0.8	[0.3;2.12]	0.66	74;40	1.04	[0.65;1.67]	0.86	44;9	NA		
Diabetes with oral medication[3]	yes	120;19	Ref.			278;129	Ref.			120;19	Ref.		
	no	80;8	2.75	[1.04;7.26]	0.04	60;42	0.73	[0.46;1.18]	0.2	80;8	NA		
Diabetes with insulin[3]	yes	123;6	Ref.			220;73	Ref.			123;6	Ref.		
	no	77;21	0.2	[0.07;0.55]	0	118;98	0.52	[0.35;0.78]	0	77;21	NA		
Diabetes age at diagnosis[3]	<55	32;5	Ref.			130;77	Ref.			32;5	Ref.		
	55-65	56;7	2.02	[0.17;24.06]	0.58	117;50	1.81	[1.06;3.08]	0.03	56;7	NA		
	>65	112;15	10.48	[0.41;270.43]	0.16	91;44	3.15	[1.74;5.69]	0	112;15	NA		

Educational level	<5 y	61;9	Ref.			74;40	Ref.			61;9	Ref.		
	6 to 9 y	70;16	0.39	[0.11;1.4]	0.15	78;60	0.57	[0.32;1.01]	0.06	70;16	0.49	[0.19;1.28]	0.15
	10 to 13 y	70;13	0.78	[0.18;3.34]	0.74	120;44	0.97	[0.51;1.84]	0.93	70;13	0.65	[0.23;1.85]	0.42
	≥14 y	63;8	1.52	[0.29;7.96]	0.62	66;27	1.17	[0.59;2.30]	0.65	63;4	0.92	[0.29;2.87]	
BMI 2 years before[2]	normal	45;2	Ref.			79;49	Ref.			66;8	Ref.		
	over	109;17	0.2	[0.04;0.99]	0.05	165;80	1.59	[0.97;2.62]	0.07	141;24	0.57	[0.24;1.4]	0.22
	obese	46;8	0.17	[0.03;0.94]	0.04	94;42	1.62	[0.93;2.84]	0.09	57;14	0.38	[0.14;1.03]	0.06
BMI at age 20	normal	163;21	Ref.			274;147	Ref.			218;37	Ref.		
	over	31;6	0.64	[0.22;1.88]	0.42	53;19	1.48	[0.81;2.69]	0.2	38;7	0.79	[0.33;1.92]	0.61
	obese	6;0	NA		NA 1	11;5	1.22	[0.38;3.89]	0.74	8;2	0.94	[0.10;8.55]	0.96
BMI at age 50	normal	39;9	Ref.			66;44	Ref.			60;17	Ref.		
	over	100;8	3.62	[1.12;11.75]	0.03	164;77	1.29	[0.78;2.13]	0.33	129;13	2.48	[1.07;5.74]	0.03
	obese	61;10	1.28	[0.43;3.79]	0.66	108;50	1.09	[0.62;1.89]	0.77	75;16	1	[0.44;2.29]	0.99
Weight gain >5kg	no	26;1	Ref.			39;19	Ref.			39;7	Ref.		
	yes	174;26	0.24	[0.03;1.95]	0.18	299;152	0.9	[0.48;1.68]	0.74	225;39	0.97	[0.39;2.39]	0.94
Weight gain >10kg	no	62;10	Ref.			102;53	Ref.			90;20	Ref.		
	yes	138;17	1.17	[0.47;2.89]	0.74	236;118	0.81	[0.53;1.25]	0.35	174;26	1.27	[0.65;2.47]	0.49
Weight loss since age 50	yes	135;21	Ref.			217;116	Ref.			175;36	Ref.		
	no	65;6	0.4	[0.13;1.21]	0.1	121;55	0.72	[0.45;1.15]	0.17	89;10	0.43	[0.2;0.96]	0.04

[1]Odds ratios (ORs) adjusted for age (<55, 55-65, 65-75, ≥75 years), gender, country, pack years (never-smokers and tertiles of pack-years) and BMI (<25, 25-30, ≥30 kg/m^2), except: sex in analyses stratified by sex, BMI in analyses stratified by obesity and pack-years in analyses stratified by smoking status.
[2]Obesity status defined based on BMI 2 years before recruitment.
[3]The association with PC risk could not be evaluated in reclassified NODM for diabetes-related variables due to lack of information on these variables.

Supplemental Table 10. Estimates for the observational and causal association between T2DM and PC and viceversa, applying different MRA methods, conducted among 1,162 cases and 752 controls with epidemiological and genetic data. T2DM status based on self-reported (SR) data.

	SR-based classification of T2DM status					SR-based classification of T2DM status				
	LSDM (N=289)[1]		NODM (N=136)[1]			LSDM (N=289)[1]		NODM (N=136)[1]	
Diabetes-->PC	**OR [95%CI][2]**	***p-value***	**OR [95%CI][2]**	***p-value***	**PC-->Diabetes**	**OR [95%CI][2]**	***p-value***	**OR [95%CI][2]**	***p-value***
Observational association study					**Observational association study**				
T2DM and PC	1.43 [1.09;1.88]	0.011	6.10 [3.45;10.8]	5.40E-13	PC and T2DM	1.45 [1.10;1.91]	0.008	6.08 [3.44;10.7]	4.80E-10
T2DM-allele score[3] and T2DM in controls	1.15 [1.09;1.21]	1.20E-05	1.31 [1.15;1.47]	0.0007	PC-allele score[4] and PC (without T2DM)	1.10 [1.06;1.13]	9.30E-09	1.10 [1.06;1.14]	4.10E-09
T2DM-allele score[3] and PC	1.01 [0.98;1.04]	0.5	1.02 [0.98;1.05]	0.315	PC-allele score[4] and T2DM	1.03 [0.99;1.06]	0.121	1.09 [1.04;1.14]	0.0002
Causal estimates: MR study					**Causal estimates: MR study**				
MRA_Wald	1.08 [0.86;1.29]	0.5	1.06 [0.95;1.17]	0.315	MRA_Wald	1.32 [0.97;1.67]	0.121	2.52 [2.07;3.03]	0.0002
TSLS Estimates	1.08 [0,87;1.33]	0.5	1.07 [0.96;1.20]	0.239	TSLS Estimates	1.31 [0.97;1.66]	0.123	2.52 [2.18;2.88]	0.0002
Inverse-variance weighted method (IVW)	0.90 [0.77;1.07]	0.238	0.98 [0.96;1.02]	0.692	Inverse-variance weighted method (IVW)	1.12 [0.89;1.41]	0.326	1.57 [1.13;2.12]	0.007
Mr-Egger regression	0.98 [0.74;1.30]	0.918	0.98 [0.96;1.02]	0.694	Mr-Egger regression	1.09 [0.78;1.51]	0.614	0.94 [0.57;1.55]	0.804
Mr-Egger Intercept	-0.022 (0.030)	0.468	0.006 (0.018)	0.964	Mr-Egger Intercept	0.007 (0.028)	0.807	0.090 (0.042)	0.027
Weighted median	0.94 [0.73;1.22]	0.649	0.99 [0.98;1.02]	0.905	Weighted median	1.18 [0.85;1.65]	0.323	1.36 [0.85;2.16]	0.197
Simple median	0.83 [0.60;1.06]	0.117	1.03 [0.94;1.13]	0.496	Simple median	1.50 [1.00;2.26]	0.049	2.37 [1.78;2.96]	<0.001

[1] LSDM and NODM was evaluated in comparison to non-diabetics (1,489 subjects: 851 PC cases and 638 controls), with subjects classified as either NODM (N=136) or LSDM (N=289) being removed, respectively.
[2] All estimates were adjusted for age (<55, 55-65, 65-75, ≥75 years), gender, BMI (<25, 25-30, ≥30 kg/m2), smoking (never-smokers and tertiles of pack-years), country and the first five principal components for population ancestries.
[3] From the 57 T2DM-SNPs, 16 were excluded due to high LD with other SNPs (r2>0.8) (Supplementary Table 1), and 6 SNPs were excluded due to their association with BMI (rs10830963, rs4430796) and smoking (rs2641348, rs13234407, rs1111875, rs2334499). The allele score, as instrumental variable, included the remaining 35 SNPs.
[4] From the 40 PC-SNPs, 2 were excluded due to high LD (r2>0.8) (Supplementary Table 1), and 5 SNPs were excluded due to their association with BMI (rs1747924.64538961, rs2816938.199985368, rs2736098.1294086, rs17688601.40866663) and smoking (rs6537481.148396094). The allele score, as instrumental variable, included the remaining 33 SNPs.

Supplemental Table 11. Estimates for the causal association between T2DM and PC and viceversa, applying different Multivariable MRA methods, conducted among 1,162 cases and 752 controls with epidemiological and genetic data. T2DM status based on self-reported (SR) data.

	SR-based classification of T2DM status			
	Single MRA		Multivariable MRA	
	OR [95%CI][1]	OR [95%CI][1]	OR [95%CI][1]	OR [95%CI][1]
Diabetes-->PC	**LSDM (N=289)[2]**	**NODM (N=136)[2]**	**LSDM[2] (X1) + BMI (X2) -> PC**	**NODM[2] (X1) + BMI (X2) -> PC**
TSLS Estimates	1.08 [0,87;1.33]	1.07 [0.96;1.20]	1.08 [0.84;1.41]	1.06 [0.95;1.19]
Inverse-variance weighted method (IVW)	0.90 [0.77;1.07]	0.98 [0.96;1.02]	0.94 [0.80;1.10]	1.00 [0.99;1.01]
Mr-Egger regression	0.98 [0.74;1.30]	0.98 [0.96;1.02]	1.00 [0.76;1.30]	1.00 [0.99;1.01]
Mr-Egger Intercept	´-0.022; p=0.468	0.006; p=0.964	´-0.014; p=0.54	0.001; p=0.93
PC-->Diabetes			**PC (X1) + BMI (X2) -> LSDM**	**PC (X1) + BMI (X2) -> NODM**
TSLS Estimates	1.03 [0.99;1.06]	2.52 [2.05;3.03]	1.05 [0.90;1.20]	1.31 [1.10;1.52]
Inverse-variance weighted method (IVW)	1.12 [0.89;1.41]	1.57 [1.13;2.12]	1.20 [0.98;1.48]	1.58 [1.15;2.17]
Mr-Egger regression	1.09 [0.78;1.51]	0.94 [0.57;1.55]	1.21 [0.87;1.69]	1.19 [0.72;1.96]
Mr-Egger Intercept	0.007; p=0.81	0.090; p=0.027	´-0.002; p=0.93	0.063; p=0.14
PC-->Diabetes (without outliers)[3]			**PC (X1) + BMI (X2) -> LSDM**	**PC (X1) + BMI (X2) -> NODM**
TSLS Estimates	1.38 [0.95;1.99]	2.85 [2.04;3.98]		
Inverse-variance weighted method (IVW)	1.18 [0.93;1.51]	1.52 [1.08;2.13]	1.19 [0.97;1.42]	1.77 [1.46;2.08]
Mr-Egger regression	1.18 [0.84;1.66]	1.36 [0.80;2.32]	1.28 [0.87;1.87]	1.65 [0.98;2.77]
Mr-Egger Intercept	0.001; p=0.96	0.023; p=0.6	´-0.01; p=0.645	0.01; p=0.73

From the 57 T2DM-SNPs, 16 were excluded due to high LD with other SNPs (r2>0.8) (Supplementary Table 1), and 4 SNPs was excluded due to its association with smoking (rs2641348, rs13234407, rs1111875, rs2334499). The allele score, as instrumental variable, included the remaining 37 SNPs. SNPs associated with obesity were not excluded.

From the 40 PC-SNPs, 2 were excluded due to high LD (r2>0.8) (Supplementary Table 1), and 1 SNPs was excluded due to its association with smoking (rs6537481.148396094). The allele score, as instrumental variable, included the remaining 37 SNPs. SNPs associated with obesity were not excluded.

[1] All estimates were adjusted for age (<55, 55-65, 65-75, ≥75 years), gender, BMI (<25, 25-30, ≥30 kg/m2), smoking (never-smokers and tertiles of pack-years), country and the first five principal components for population ancestries.

[2] LSDM and NODM was evaluated in comparison to non-diabetics (1,489 subjects: 851 PC cases and 638 controls), with subjects classified as either NODM (N=136) or LSDM (N=289) being removed, respectively. BMI to define obesity (yes, no) two years before recruitment.

[3] Outliers removed: "rs1747924:64538961:C:A", "rs1486134:67639769:G:T", "rs17688601:40866663:C:A" for LDSM and "rs6971499" and "rs7310409" for NODM

Supplemental Table 12. Estimates for the observational and causal association between obesity measures and PC, applying different MRA methods, conducted among 1,162 cases and 752 controls with epidemiological and genetic data.

	SR-based classification of obesity status			
	BMI 2 years (N=343 obese)		BMI 50 years (N=401 obese)	
BMI-->PC	**OR [95%CI][1]**	***p-value***	**OR [95%CI][1]**	***p-value***
Observational association study				
BMI and PC	0.89 [0.69;1.14]	0.356	0.83 [0.64;1.07]	1.41E-01
BMI-allele score[2] and BMI in controls	1.11 [1.08;1.15]	3.06E-09	1.16 [1.12;1.21]	1.71E-11
BMI-allele score[2] and PC	1.01 [0.98;1.02]	0.927	1.01 [0.99;1.03]	0.311
Causal estimates: MR study				
MRA_Wald	1.01 [0.98;1.03]	0.927	1.07 [0.94;1.21]	0.311
TSLS Estimates	1.01 [0,85;1.20]	0.927	1.09 [0.95;1.25]	0.293
Inverse-variance weighted method (IVW)	1.01 [0.90;1.15]	0.828	1.03 [0.93;1.14]	0.542
Mr-Egger regression	0.98 [0.74;1.30]	0.794	0.97 [0.82;1.15]	0.705
Mr-Egger Intercept	0.009(0.015)	0.573	0.016 (0.017)	0.343
Weighted median	0.96 [0.80;1.32]	0.688	1.03 [0.88;1.20]	0.728
Simple median	1.10 [0.92;1.32]	0.278	1.05 [0.90;1.22]	0.554

[1] All estimates were adjusted for age (<55, 55-65, 65-75, ≥75 years), gender, smoking (never-smokers and tertiles of pack-years), country and the first five principal components for population ancestries.
[4] From the 126 obesity-SNPs, 30 were excluded due to high LD (r2>0.8) (Supplementary Table 1), and 11 SNPs were excluded due to their association with T2DM and smoking. The allele score, as instrumental variable, included the remaining 85 SNPs.
There were few obese subjects at age 20 years; BMI at this age was therefore not considered.

Supplemental Table 13. Results of causal mediation analyses evaluating mediator effects of T2DM on the obesity and PC association considering different obesity measures, and mediator effects of obesity on the T2DM and PC association. Estimates are derived from counterfactual models (2,018 cases and 1,540 controls) and MRA (1,162 cases and 752 controls with epidemiological and genetic data).

	OR[1]	[95%CI]	OR[1]	[95%CI]	OR[1]	[95%CI]	OR[1]	[095%CI]	OR[1]	[95%CI]	OR[1]	[95%CI]
Counterfactual model	**NDE**		**NIE**		**TE**		**NDE**		**NIE**		**TE**	
NODM mediator							**LSDM mediator**					
Obese[2]	0.90	[0.73;1.10]	**1.04**	**[1.01;1,11]**	0.94	[0.77;1.17]	0.95	[0.80;1.15]	**1.05**	**[1.02;1.10]**	1.00	[0.84;1.22]
Overweight/obese[2]	**0.83**	**[0.71;0.98]**	**1.09**	**[1.08;1,13]**	0.91	[0.79;1.11]	0.92	[0.80;1.08]	**1.03**	**[1.01;1.06]**	0.95	[0.82;1.11]
Weight gain > 5 kg[3]	**0.82**	**[0.66;0.97]**	**1.08**	**[1.07;1.10]**	0.89	[0.72;1.07]	0.86	[0.70;1.03]	**1.05**	**[1.03;1.08]**	0.89	[0.74;1.08]
Weight loss[3]	1.04	[0.85;1.38]	0.95	[0.73;1.02]	1.00	[0.83;1.19]	0.98	[0.82;1.15]	**1.02**	**[1.00;1.04]**	1.00	[0.84;1.17]
Obese at age 50[3]	0.83	[0.67;1.01]	**1.07**	**[1.04;1,13]**	0.89	[0.73;1.08]	0.88	[0.72;1.07]	**1.07**	**[1.04;1.15]**	0.94	[0.78;1.14]
Overweight/obese at age 50[3]	**1.21**	**[1.01;1.44]**	1.03	[0.98;1.08]	**1.25**	**[1.07;1.51]**	**1.18**	**[1.01;1.37]**	**1.04**	**[1.03;1.07]**	**1.23**	**[1.06;1.43]**
Obese[2] mediator							**Overweight/obese[2] mediator**					
NODM	**5.92**	**[3.69;9.14]**	0.97	[0.78;1.01]	**5.72**	**[3.76;9.11]**	**10.14**	**[5.48;22.69]**	**0.55**	**[0.23;0.92]**	**5.58**	**[3.65;8.92]**
LSDM	**1.65**	**[1.34;2.03]**	1.02	[0.99;1.07]	**1.68**	**[1.37;2.06]**	**1.61**	**[1.31;2.00]**	**1.03**	**[1.01;1.08]**	**1.67**	**[1.35;2.06]**
Obese at age 50[3] mediator							**Overweight/obese at age 50[3] mediator**					
NODM	**4.99**	**[2.54;10.87]**	0.87	[0.43;1.04]	**4.35**	**[2.35;9.65]**	**4.27**	**[2.14;8.08]**	**1.08**	**[1.03;1.12]**	**4.63**	**[2.66;10.82]**
LSDM	**1.53**	**[1.13;2.09]**	0.99	[0.87;1.07]	**1.50**	**[1.14;2.17]**	**1.38**	**[1.08;1.94]**	**1.09**	**[1.03;1.15]**	**1.49**	**[1.17;2.02]**
Counterfactual IV	**NDE**		**NIE**		**TE**		**NDE**		**NIE**		**TE**	
NODM mediator							**LSDM mediator**					
Obese[2]	0.83	[0.63;1.15]	1.03	[1.00;1.11]	0.85	[0.68;1.24]	0.88	[0.63;1.27]	1.02	[0.99;1.04]	0.89	[0.68;1.27]
Overweight/obese[2]	0.93	[0.72;1.12]	1.05	[1.00;1.08]	0.97	[0.81;1.26]	0.98	[0.81;1.27]	1.01	[0.99;1.02]	0.98	[0.79;1.27]
Weight gain > 5 kg[3]	1.15	[0.68;1.53]	1.03	[1.00;1.07]	1.19	[0.68;1.56]	1.16	[0.94;1.58]	1.02	[1.00;1.03]	1.18	[0.95;1.78]
Weight loss[3]	1.20	[0.96;1.69]	**0.94**	**[0.90;0.98]**	1.13	[0.92;1.48]	1.13	[0.88;1.35]	1.01	[0.99;1.02]	1.14	[0.88;1.39]
Obese at age 50[3]	0.82	[0.66;1.22]	1.06	[1.00;1.12]	0.87	[0.70;1.31]	0.84	[0.66;1.12]	1.02	[1.00;1.06]	0.86	[0.68;1.13]
Overweight/obese at age 50[3]	**1.44**	**[1.18;1.89]**	1.03	[1.00;1.06]	**1.49**	**[1.23;1.96]**	**1.47**	**[1.17;1.88]**	**1.02**	**[1.01;1.03]**	**1.49**	**[1.19;1.90]**
Obese[2] mediator							**Overweight/obese[2] mediator**					
NODM	**5.06**	**[3.15;13.19]**	0.92	[0.40;1.01]	**4.67**	**[2.97;9.77]**	**8.21**	**[6.64;11.58]**	**0.53**	**[0.38;0.75]**	**4.37**	**[3.02;7.75]**
LSDM	**1.46**	**[1.02;1.91]**	1.02	[0.97;1.07]	**1.47**	**[1.07;2.01]**	**1.45**	**[1.12;2.18]**	1.01	[0.94;1.04]	**1.46**	**[1.22;2.29]**
Obese at age 50[3] mediator							**Overweight/obese at age 50[3] mediator**					
NODM	**5.01**	**[3.02;12.20]**	0.86	[0.37;1.06]	**4.32**	**[2.78;8.77]**	**4.28**	**[2.72;9.79]**	**1.09**	**[1.05;1.20]**	**4.68**	**[3.04;11.49]**
LSDM	**1.51**	**[1.24;2.20]**	0.98	[0.87;1.04]	**1.49**	**[1.25;2.32]**	**1.37**	**[0.88;1.81]**	**1.08**	**[1.03;1.16]**	**1.48**	**[1.07;2.15]**

CI, confidence interval; TE, marginal total effect; NDE, natural direct effect; NIE, natural indirect effect;
[1] All estimates were adjusted for age (<55, 55-65, 65-75, ≥75 years), gender, smoking (never-smokers and tertiles of pack-years), country, and the first five principal components for population ancestries in network MRA
[2] Obesity status defined based on BMI 2 years before recruitment
[3] Obesity-related variables based on information collected at age 50 years, such as weight gain from age 20 to 50 and weight loss since age 50 years
Significant estimates are marked in bold.

Supplemental Table 14. Association between diabetes-related variables and PC risk in the PanGenEU study (2,018 cases and 1,540 controls). Unimputed data.

	Cases N=2,018		Controls N=1,540			Model1		Model2	
	N	%	N	%	p-value[1]	OR	[95%CI]	OR	[95%CI]
Diabetes status					<0.001				
no diabetes	1479	73.30	1340	87.00		Ref.		Ref.	
yes	498	24.70	184	11.90		2.60	[2.13;3.18]	2.55	[2.04;3.18]
Missing	41	2.03	16	1.04					
Diabetes status by subtype					<0.001				
no diabetes	1479	73.30	1340	87.00		Ref.		Ref.	
yes, ≤ 2 years (NODM)	200	9.91	27	1.75		6.41	[4.2;9.79]	6.43	[4.06;10.2]
yes, > 2years (LSDM)	265	13.10	152	9.87		1.82	[1.45;2.3]	1.77	[1.37;2.28]
Missing	74	3.67	21	1.36					
Family history of diabetes[2]									
no diabetes	1069	58.0	821	65.7		Ref.		Ref.	
yes	594	32.2	357	28.6		1.24	[1.04;1.49]	1.14	[1.00;1.38]
Missing	182	9.9	72	5.8					
Diabetes by age at diagnosis[3]					<0.001				
no diabetes	1479	73.30	1340	87.00		Ref.		Ref.	
≤ 55 years	141	6.99	72	4.68		1.66	[1.21;2.28]	1.59	[1.12;2.26]
55 to ≤ 65 years	138	6.84	50	3.25		2.48	[1.74;3.55]	2.48	[1.69;3.64]
> 65 years	197	9.76	58	3.77		3.97	[2.88;5.54]	3.79	[2.67;5.44]
Missing	63	3.12	20	1.30	*p-trend*	2E-16		2E-16	
Diabetes by time since diagnosis[3]					<0.001				
no diabetes	1479	73.30	1340	87.0		Ref.		Ref.	
≤1	159	7.88	12	0.78		10.98	[6;20.09]	9.39	[5.08;17.34]
1 to ≤2	41	2.03	15	0.97		2.64	[1.4;4.97]	3.19	[1.56;6.52]
2 to ≤5	71	3.52	32	2.08		2.38	[1.54;3.75]	2.43	[1.52;3.94]
5 to ≤10	86	4.26	36	2.34		2.43	[1.60;3.73]	2.41	[1.52;3.86]
10 to ≤20	56	2.78	40	2.60		1.50	[0.96;2.33]	1.38	[0.85;2.25]
>20	52	2.58	44	2.86		1.21	[0.78;1.90]	1.11	[0.67;1.81]
Missing	74	3.67	21	1.36	*p-trend*	1.5E-06		4.5E-05	
Diabetes control measures									
Diet					<0.001				
no diabetes	1479	73.30	1340	87.00		Ref.		Ref.	
yes	297	14.70	133	8.64		2.50	[1.98;3.15]	2.44	[1.9;3.14]
no use	118	5.85	49	3.18		2.41	[1.68;3.46]	2.38	[1.6;3.55]
Missing	124	6.14	18	1.17					
Use of oral medication					<0.001				
no diabetes	1479	73.30	1340	87.00		Ref.		Ref.	
yes	304	15.10	130	8.44		2.36	[1.86;2.99]	2.24	[1.73;2.9]
no use	140	6.94	50	3.25		3.00	[2.12;4.24]	3.21	[2.17;4.74]
Missing	95	4.71	20	1.30					
Use of insulin					<0.001				
no diabetes	1479	73.30	1340	87.00		Ref.		Ref.	
yes	236	11.70	59	3.83		4.26	[3.12;5.81]	4.18	[2.97;5.89]
no use	195	9.66	119	7.73		1.77	[1.37;2.28]	1.77	[1.34;2.35]
Missing	108	5.35	22	1.43					

[1] Differences between groups evaluated by the Chi-square test
[2] Information on family history of diabetes was not collected in Ireland; results are based on data for 1,845 cases and 1,250 controls
[3] Linear association for age since T2DM diagnosis and nonlinear association for time since T2DM (Supplemental Figure 1) Model 1: adjusted for age (<55, 55-65, 65-75, ≥75 years), gender, country. Model 2: Model 1 also adjusted for pack years (never-smokers and tertiles of pack-years) and BMI (<25, 25-30, ≥30 kg/m^2)

Supplemental Table 15. Estimates for the observational and causal association between T2DM and PC and vice versa, applying different MRA methods, conducted among 1,162 cases and 752 controls with epidemiological and genetic data. T2DM status based on self-reported (SR) and biomarker data.

	SR + biomarker-based classification of T2DM status					SR + biomarker -based classification of T2DM status				
	LSDM (N=289)[1]		NODM (N=190)[1]			LSDM (N=289)[1]		NODM (N=190)[1]		
Diabetes-->PC	**OR [95%CI][2]**	***p-value***	**OR [95%CI][2]**	***p-value***	**PC-->Diabetes**	**OR [95%CI][2]**	***p-value***	**OR [95%CI][2]**	***p-value***
Observational association study					**Observational association study**				
T2DM and PC	1.50 [1.14;1.98]	0.003	5.08 [3.27;7.90]	4.40E-13	PC and T2DM	1.51 [1.15;2.00]	0.003	5.15 [3.31;8.00]	3.22E-13
T2DM-allele score[3] and T2DM in controls	1.11 [1.05;1.16]	3.73E-04	1.23 [1.13;1.33]	4.74E-05	PC-allele score[4] and PC (without T2DM)	1.10 [0.75;1.45]	1.54E-08	1.09 [1.07;1.13]	1.54E-08
T2DM-allele score[3] and PC	1.02 [0.99;1.05]	0.146	0.99 [0.96;1.02]	0.461	PC-allele score[4] and T2DM	1.03 [0.99;1.06]	0.119	1.07 [1.03;1.11]	0.0014
Causal estimates: MR study					**Causal estimates: MR study**				
MRA_Wald	1.21 [0.95;1.47]	0.146	0.95 [0.96;1.02]	0.461	MRA_Wald	1.32 [0.97;1.67]	0.12	2.01 [1.58;2.43]	0.0014
T2LS Estimates	1.19 [0.92;1.54]	0.194	0.95 [0.84;1.08]	0.461	T2LS Estimates	1.03 [0.99;1.06]	0.12	2.86 [2.07;3.97]	2.37E-10
Inverse-variance weighted method (IVW)	1.06 [0.79;1.42]	0.708	0.99 [0.93;1.05]	0.725	Inverse-variance weighted method (IVW)	1.12 [0.89;1.41]	0.316	1.29 [0.98;1.70]	0.078
Mr-Egger regression	1.31 [0.80;2.15]	0.278	1.00 [0.94;1.06]	0.921	Mr-Egger regression	1.05 [0.76;1.47]	0.756	0.83 [0.55;1.26]	0.382
Mr-Egger Intercept	´-0.049 (0.045)	0.283	´-0.019 (0.029)	0.538	Mr-Egger Intercept	0.015 (0.028)	0.604	0.095 (0.034)	0.005
Weighted median	0.97 [0.72;1.30]	0.823	1.00 [0.96;1.04]	0.812	Weighted median	1.22 [0.88;1.17*	0.238	1.16 [0.78;1.75]	0.446
Simple median	1.20 [0.86;1.70]	0.284	1.00 [0.85;1.17]	0.95	Simple median	1.55 [1.00;2.32]	0.05	1.56 [0.96;2.54]	0.075

[1] LSDM and NODM was evaluated in comparison to non-diabetics after reclassifying T2DM status with the biomarker data (obtained for 654 subjects with epidemiological and genetic data), with subjects reclassified as either NODM (N=190) or LSDM (N=289) being removed, respectively.
[2] All estimates were adjusted for age (<55, 55-65, 65-75, ≥75 years), gender, BMI (<25, 25-30, ≥30 kg/m2), smoking (never-smokers and tertiles of pack-years), country and the first five principal components for population ancestries.
[3] From the 57 T2DM-SNPs, 16 were excluded due to high LD with other SNPs (r2>0.8) (Supplementary Table 1), and 6 SNPs were excluded due to their association with BMI (rs10830963, rs4430796) and smoking (rs2641348, rs13234407, rs1111875, rs2334499). The allele score, as instrumental variable, included the remaining 35 SNPs.
[4] From the 40 PC-SNPs, 2 were excluded due to high LD (r2>0.8), and 5 SNPs were excluded due to their association with BMI (rs1747924.64538961, rs2816938.199985368, rs2736098.1294086, rs17688601.40866663) and smoking (rs6537481.148396094). The allele score, as instrumental variable, included the remaining 33 SNPs.

Removal of SNPs potentially associated with other traits (at p-value 10-8) according to PhenoScanner database led to similar results.

Supplemental Table 16. Estimates for the observational and causal association between T2DM and PC and vice versa, after removing other potential pleiotropic variants and outliers (based on Cook´s distances) and applying different MRA methods, conducted among 1,162 cases and 752 controls with epidemiological and genetic data. T2DM status based on self-reported (SR) data.

	SR-based classification of T2DM status					SR-based classification of T2DM status				
	LSDM (N=289)[1]		NODM (N=136)[1]			LSDM (N=289)[1]		NODM (N=136)[1]		
Diabetes-->PC	**OR [95%CI][2]**	***p-value***	**OR [95%CI][2]**	***p-value***	**PC-->Diabetes**	**OR [95%CI][2]**	***p-value***	**OR [95%CI][2]**	***p-value***
Observational association study					**Observational association study**				
T2DM and PC	1.43 [1.09;1.88]	0.011	6.10 [3.45;10.8]	<0.001	PC and T2DM	1.45 [1.10;1.91]	0.008	6.08 [3.44;10.7]	4.80E-10
T2DM-allele score[3] and T2DM in controls	1.16 [1.09;1.22]	1.42E-05	1.32 [1.15;1.48]	0.001	PC-allele score[4] and PC (without T2DM)	1.10 [1.06;1.13]	4.10E-08	1.10 [1.06;1.13]	3.10E-08
T2DM-allele score[3] and PC	1.00 [0.97;1.03]	9.93E-01	1.01 [0.98;1.05]	0.389	PC-allele score[4] and T2DM	1.03 [1.00;1.06]	0.09	1.08 [1.03;1.13]	0.0023
Causal estimates: MR study					**Causal estimates: MR study**				
MRA_Wald	1.00 [0.79;1.22]	0.273	1.05 [0.99;1.09]	0.389	MRA_Wald	1.38 [0.99;1.74]	0.09	2.47 [1.71;2.74]	0.00053
TSLS Estimates	0.98 [0.79;1.22]	0.864	1.05 [0.93;1.19]	0.39	TSLS Estimates	1.38 [0.95;1.99]	0.09	2.85 [2.04;3.98]	2.80E-09
Inverse-variance weighted method (IVW)	0.92 [0.79;1.08]	0.315	1.00 [0.96;1.02]	0.538	Inverse-variance weighted method (IVW)	1.18 [0.93;1.51]	0.16	1.52 [1.08;2.13]	0.016
Mr-Egger regression	1.00 [0.77;1.20]	0.992	0.98 [0.96;1.02]	0.534	Mr-Egger regression	1.18 [0.84;1.66]	0.348	1.36 [0.80;2.32]	0.251
Mr-Egger Intercept	0.019 (0.025)	0.448	0.001 (0.018)	0.936	Mr-Egger Intercept	0.001 (0.028)	0.963	0.023 (0.043)	0.06
Weighted median	0.96 [0.74;1.24]	0.481	1.01 [0.95;1.04]	0.966	Weighted median	1.18 [0.84;1.67]	0.321	1.43 [0.88;2.33]	0.15
Simple median	0.86 [0.69;1.07]	0.174	1.02 [0.92;1.11]	0.735	Simple median	1.51 [1.00;2.27]	0.052	2.84 [2.27;3.41]	<0.001

[1] LSDM and NODM was evaluated in comparison to non-diabetics (1,489 subjects: 851 PC cases and 638 controls), with subjects classified as either NODM (N=136) or LSDM (N=289) being removed, respectively.
[2] All estimates were adjusted for age (<55, 55-65, 65-75, ≥75 years), gender, BMI (<25, 25-30, ≥30 kg/m2), smoking (never-smokers and tertiles of pack-years), country and the first five principal components for population ancestries.
[3] From the 57 T2DM-SNPs, 16 were excluded due to high LD with other SNPs (r2>0.8) (Supplementary Table 1), and 6 SNPs were excluded due to their association with BMI (rs10830963, rs4430796) and smoking (rs2641348, rs13234407, rs1111875, rs2334499). In addition, 3 SNPs potentially being outliers were removed (rs2191348, rs13266634, rs7965349). The allele score, as instrumental variable, included the remaining 32 SNPs.
[4] From the 40 PC-SNPs, 2 were excluded due to high LD (r2>0.8) (Supplementary Table 1), and 5 SNPs were excluded due to their association with BMI (rs1747924.64538961, rs2816938.199985368, rs2736098.1294086, rs17688601.40866663) and smoking (rs6537481.148396094). In addition, 2 SNPs potentially being outliers were removed (chr12_121454622, chr16_75263661). The allele score, as instrumental variable, included the remaining 31 SNPs.

Supplemental Table 17. Magnitudes of the E-value for different combinations of the Exposure-Confounder Association RREU and the Confounder-Outcome Association RRUD for the estimation of the causal effect of NODM on PC (OR=6.39 (4.18;9.78)) and of LSDM on PC (OR=1.86 (1.49;2.32)).

NDOM

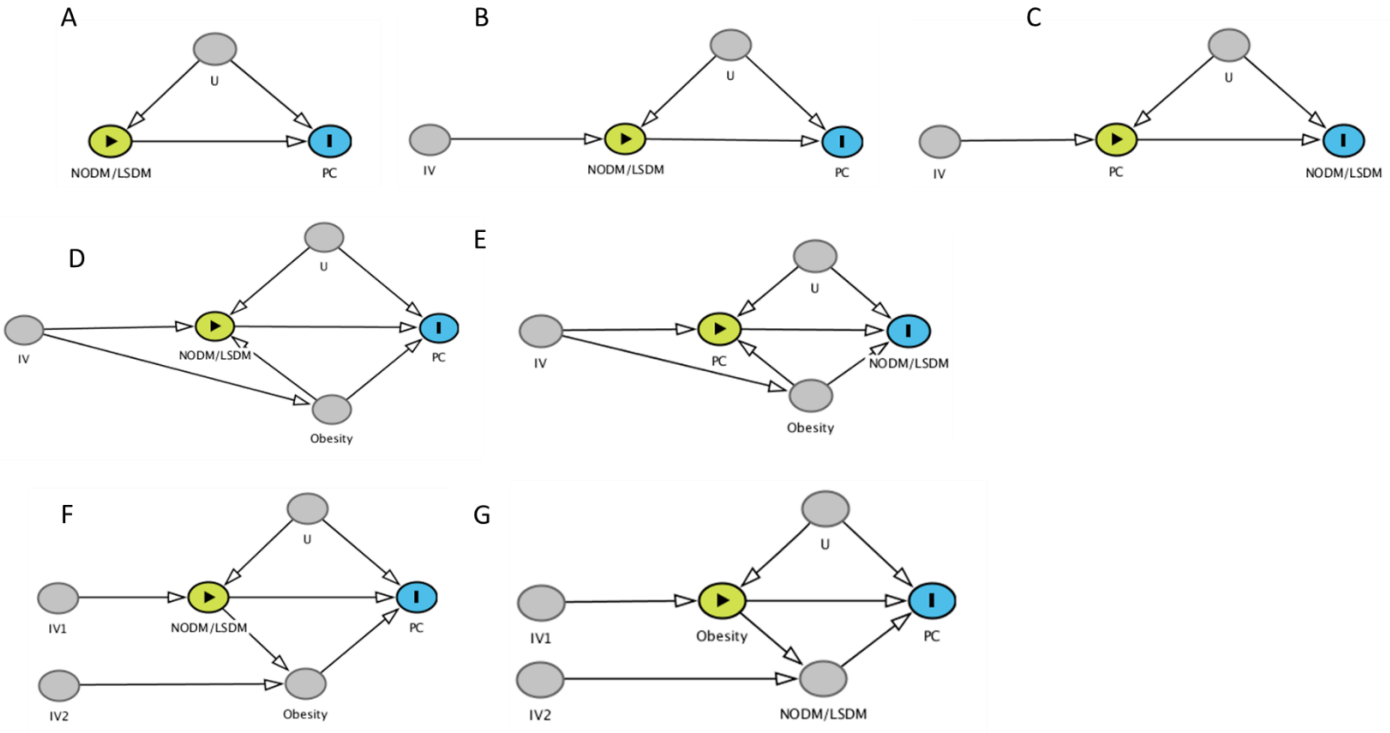
E-value		RR_{UD}								
		3.5	**6.5**	**9.5**	**12.5**	**15.5**	**18.5**	**21.5**	**24.5**	**27.5**
RR_{EU}	**3.5**	2,04	2,53	2,77	2,92	3,01	3,08	3,14	3,18	3,21
	6.5	2,53	3,52	4,12	4,51	4,80	5,01	5,18	5,31	5,42
	9.5	2,77	4,12	5,01	5,65	6,14	6,51	6,81	7,05	7,26
	12.5	2,92	4,51	5,65	6,51	7,18	7,71	8,14	8,51	8,81
	15.5	3,01	4,80	6,14	7,18	8,01	8,69	9,26	9,74	10,15
	18.5	3,08	5,01	6,51	7,71	8,69	9,51	10,20	10,79	11,31
	21.5	3,14	5,18	6,81	8,14	9,26	10,20	11,01	11,71	12,32
	24.5	3,18	5,31	7,05	8,51	9,74	10,79	11,71	12,51	13,21
	27.5	3,21	5,42	7,26	8,81	10,15	11,31	12,32	13,21	14,00

LSDM

E-value		RR_{UD}									
		1.5	**2.5**	**3.5**	**4.5**	**5.5**	**6.5**	**7.5**	**8.5**	**9.5**	**10.5**
RR_{EU}	**1.5**	1,13	1,25	1,31	1,35	1,38	1,39	1,41	1,42	1,43	1,43
	2.5	1,25	1,56	1,75	1,88	1,96	2,03	2,08	2,13	2,16	2,19
	3.5	1,31	1,75	2,04	2,25	2,41	2,53	2,63	2,70	2,77	2,83
	4.5	1,35	1,88	2,25	2,53	2,75	2,93	3,07	3,19	3,29	3,38
	5.5	1,38	1,96	2,41	2,75	3,03	3,25	3,44	3,60	3,73	3,85
	6.5	1,39	2,03	2,53	2,93	3,25	3,52	3,75	3,95	4,12	4,27
	7.5	1,41	2,08	2,63	3,07	3,44	3,75	4,02	4,25	4,45	4,63
	8.5	1,42	2,13	2,70	3,19	3,60	3,95	4,25	4,52	4,75	4,96
	9.5	1,43	2,16	2,77	3,29	3,73	4,12	4,45	4,75	5,01	5,25
	10.5	1,43	2,19	2,83	3,38	3,85	4,27	4,63	4,96	5,25	5,51

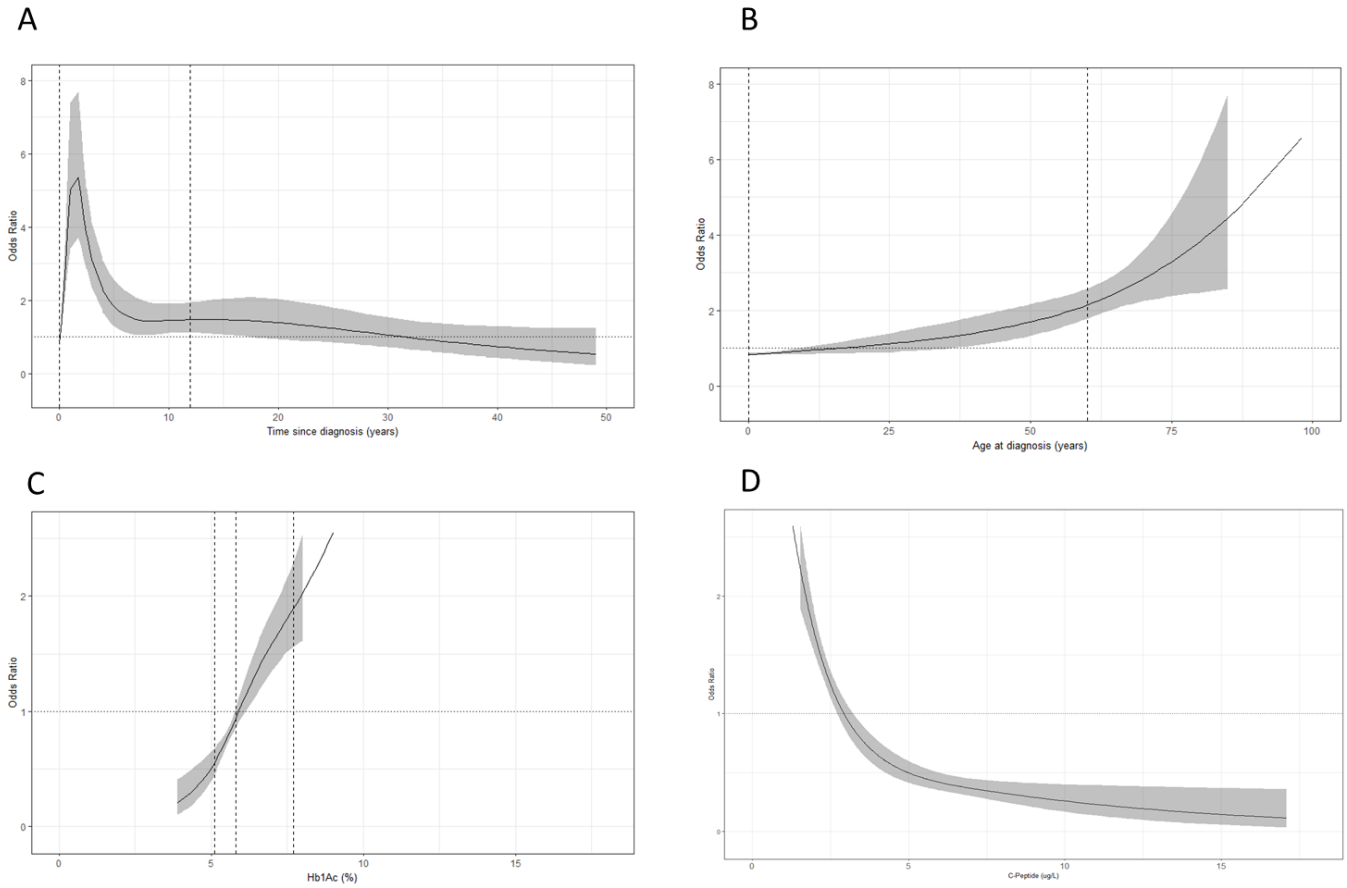
Supplementary Figures

Supplemental Figure 1. Directed acyclic graphs illustrating the single MR and multivariable and network MR approaches used to explore causal associations and mediation in the causal pathways between T2DM, obesity and PC.

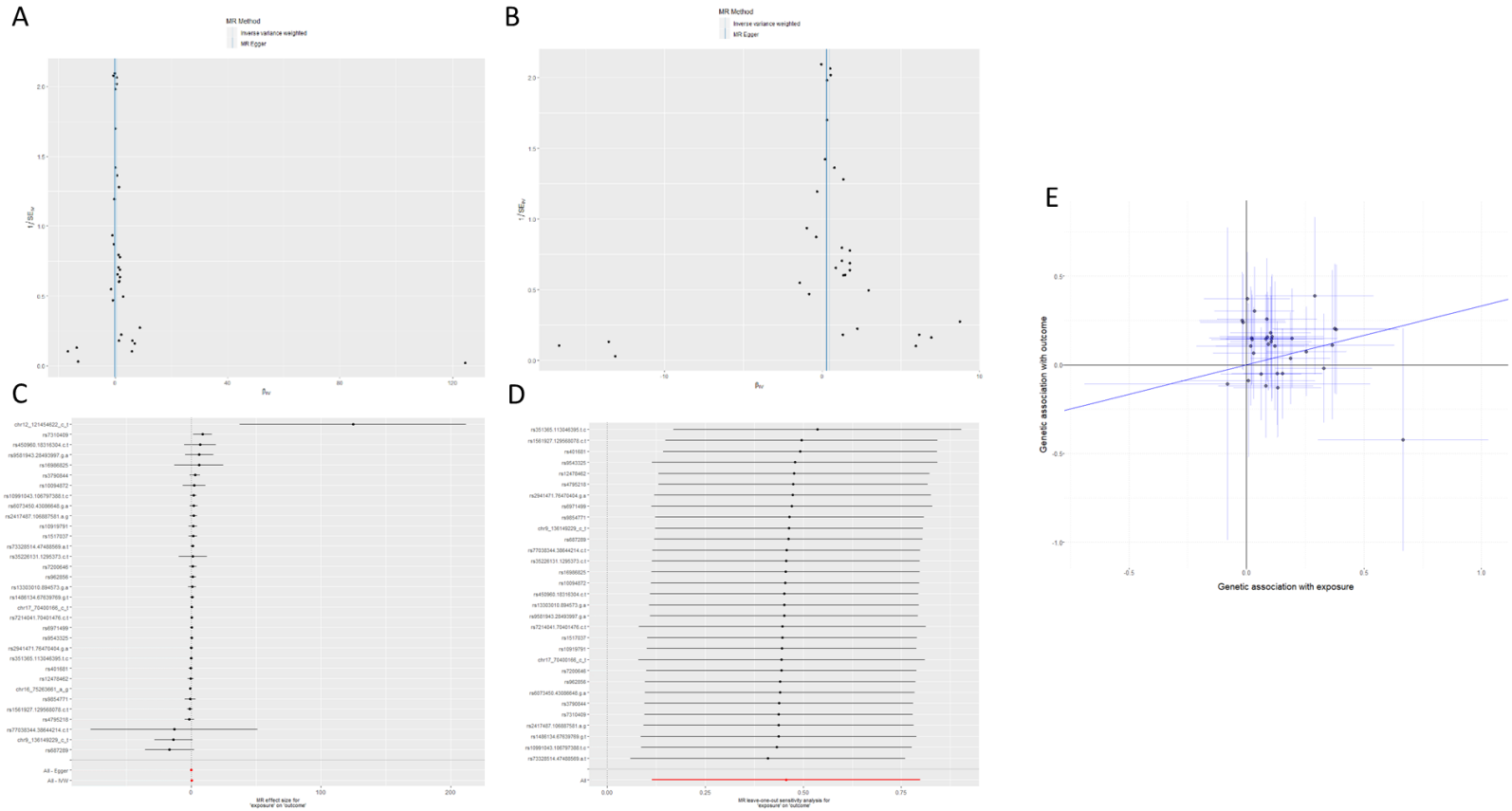


A: Observational association between T2DM (Exposure) and PC (Outcome)
B: Single Mendelian Randomization (MR) between T2DM (Exposure) and PC (Outcome)
C: Single Mendelian Randomization (MR) between PC (Exposure) and T2DM (Outcome) – bidirectional MR
D: Multivariable Mendelian Randomization (MR) between T2DM (Exposure) and PC (Outcome)
E: Multivariable Mendelian Randomization (MR) between PC (Exposure) and T2DM (Outcome) – bidirectional MR
F: Network Mendelian Randomization (MR) between T2DM (Exposure), Obesity (Mediator) and PC (Outcome)
G: Network Mendelian Randomization (MR) between Obesity (Exposure), T2DM (Mediator) and PC (Outcome)

Supplemental Figure 2. Linear and Non-linear association between T2DM-related continuous variables and pancreatic cancer risk, with non-diabetics as a reference group: (A) time since T2DM diagnosis; (B) age at T2DM diagnosis; (C) Hb1Ac levels and (D) C-Peptide levels with the minimum value (Hb1Ac=4; C-Peptide=0.05) as the reference group.



Supplemental Figure 3. Pleiotropy visualization plots regarding the directional association between PC and NODM risk.



A: funnel plot for IV made up of SNPs without SNPs in LD and SNPs associated with obesity and smoking. **B:** funnel plot for IV excluding further SNPs that were outliers (based on Cooks distances). Y-axes represent SNP to outcome effect corrected by SNP to exposure standard error of the effect. X-axes (SNP to exposure effect) are in logarithmic scale. **C:** Correlation plot of per-allele associations (genetic score of the IV) with the outcome and exposure. **D:** Forest plot of per-allele MR effect size for exposure on outcome and **E:** leave-one-out analyses.

Appendix B. PanGenEU Study

The PanGenEU (the European Study into Digestive Illnesses and Genetics) is a multicentre hospital-based case control study of pancreatic cancer conducted in six European countries (Spain, Germany, Ireland, United Kingdom, Italy and Sweden) and 28 centres, designed to evaluate environmental and genetic factor associated with pancreatic cancer risk.

Epidemiological information was obtained through an epidemiological questionnaire including self-reported socio-demographic, anthropometric data (height, weight), family history of cancer including pancreatic cancer, medical history (diabetes), lifestyle behaviours (e.g., smoking and alcohol habits).

Genetic information as SNPs (Single Nucleotide Polymorphism) was available for 1,162 cases and 540 controls who provided blood samples. DNA samples were genotyped on the Infinium OncoArray-500K at the CEGEN (Spanish National Cancer Research Centre, CNIO). The genotype data was filtered for call rate, relatedness, European ancestry <80% and sex chromosome abnormalities. Overall, 451,883 SNPs passed these quality filters and underwent imputation of missing genotypes using IMPUTE v2. The control group was enlarged with 212 controls participating in two Spanish bladder cancer case-control studies (EPICURO and ISBlaC), with analogous characteristics to the source population (Spanish pancreatic cancer cases; 44% females and mean age=64.7 years) and with genotype and epidemiological data available. Similar protocols for data collection and genotyping were used in all studies.

The construction of the Genetic Score as an Instrumental Variable was made considering SNPs that were statistically associated with the exposure of interest. The genetic score is a weighted score where each value of the SNP multiplying by the real effect (μ) of each SNP to the exposure.

$$\mathrm{IV}_1 = \sum_{j=1}^{n_1} \mu_j \cdot SNP_{BMI_j}$$

$$\mathrm{IV}_2 = \sum_{j=1}^{n_2} \mu_j \cdot SNP_{T2DM}$$

Where IV_1 is the weighted genetic score as IV for body mass index (BMI) exposure and n_1 is the number of significative SNPs associated with BMI. IV_2 is the weighted genetic score for type 2 diabetes mellitus (T2DM) and n_2 is the number of SNPs associated with T2DM. For this study, we discarded those SNPs that were associated with both BMI and T2DM and SNPs associated to each other, in order to prevent any pleiotropy problem and Linkage Disequilibrium (LD).

Appendix C. Simulation process

Based on the PanGenEU features, we create several simulation datasets considering different parameters and scenarios.

S1: *Y*, *E*, and *M* continuous

Step 1. Selecting effect sizes based on the real dataset, PanGenEU.

To select the effect sizes, we fitted the following regression models for E, body mass index, M, glycated haemoglobin, and Y, pancreatic cancer, respectively:

$$E = a_0 + a_1 Sex + a_2 Smoke[1] + a_3 Smoke[2] + a_4 Smoke[3] + a_5 Age + a_6 IV_1 + \varepsilon_E \quad [C1]$$

$$M = b_0 + b_1 Sex + b_2 Smoke[1] + b_3 Smoke[2] + b_4 Smoke[3] + b_5 Age + b_6 IV_2 + b_7 E + \varepsilon_M \quad [C2]$$

$$Y^* = c_0 + c_1 E + c_2 M + c_3 Sex + c_4 Age + c_5 Smoke[1] + c_6 Smoke[2] + c_7 Smoke[3] + \varepsilon_Y \quad [C3]$$

Y^* corresponds to the probit score of pancreatic cancer.

The estimated coefficients $\hat{a} = (\hat{a}_0, \hat{a}_1, ..., \hat{a}_6)$; $\hat{b} = (\hat{b}_0, \hat{b}_1, ..., \hat{b}_7)$; $\hat{c} = (\hat{c}_0, \hat{c}_1, ..., \hat{c}_7)$ and the estimated standard deviations of the errors $(\hat{\sigma}_E, \hat{\sigma}_M, \hat{\sigma}_Y)$ associated to E, M and Y, respectively will be used in Step 4. For the sake of simplicity, vector $(\hat{\sigma}_E, \hat{\sigma}_M, \hat{\sigma}_Y)$ has been denote $(\sigma_E, \sigma_M, \sigma_Y)$ in the whole paper.

Step 2. Simulating correlated errors.

The errors were simulated assuming a multivariate normal distribution:

$$(\varepsilon_{ESim}, \varepsilon_{MSim}, \varepsilon_{YSim}) \sim N_3(\mu, \Sigma)$$

with $\mu = (0, 0, 0)$ and Σ, the variance-covariance matrix defined as:

$$\Sigma = \begin{pmatrix} \sigma_E^2 & \rho\sigma_E\sigma_M & \rho\sigma_E\sigma_Y \\ \rho\sigma_E\sigma_M & \sigma_M^2 & \rho\sigma_M\sigma_Y \\ \rho\sigma_E\sigma_Y & \rho\sigma_M\sigma_Y & \sigma_Y^2 \end{pmatrix}$$

where $(\sigma_E, \sigma_M, \sigma_Y)$ were those obtained in Step 1.

Step 3. Selecting confounders, and instrumental variables.

Instead of simulating artificial datasets, confounders, and instrumental variables were randomly extracted from the real dataset with replacement.

Step 4. Simulating *E*, *M* and *Y*.

In the simulated datasets, the exposure, E, the mediator, M, and the outcome, Y, were obtained as follows:

Step 4.1:

$$E_{sim} = \hat{a}_0 + \hat{a}_1 Sex + \hat{a}_2 Smoke[1] + \hat{a}_3 Smoke[2] + \hat{a}_4 Smoke[3] + \hat{a}_5 Age + \hat{a}_6 IV_1 + \varepsilon_{E_{Sim}} \quad [C4]$$

Step 4.2:

$$M_{sim} = \hat{b}_0 + \hat{b}_1 Sex + \hat{b}_2 Smole[1] + \hat{b}_3 Smoke[2] + \hat{b}_4 Smoke[3] + \hat{b}_5 Age + \hat{b}_6 IV_2 + \hat{b}_7 E_{sim} + \varepsilon_{M\,sim} \quad [C5]$$

where E_{Sim} are the values obtained from [C4]

Step 4.3:

$$Y_{sim} = \hat{c}_0 + \hat{c}_1 E_{sim} + \hat{c}_2 M_{sim} + \hat{c}_3 Sex + \hat{c}_4 Age + \hat{c}_5 Smoke[1] + \hat{c}_6 Smoke[2] + \hat{c}_7 Smoke[3] + \varepsilon_{Y_{sim}} \quad [C6]$$

where E_{sim} and M_{sim} are the values obtained from [C4] and [C5], respectively.

The process for the rest of the scenarios only differs in Steps 1 and 4.

S2 : *Y* binary, *E* continuous and *M* continuous

Step 1. Selecting effect sizes based on the real dataset, PanGenEU.

As [C1], [C2] and [C3], but in this case Y^* corresponds to the probit score in SEM and logit score in *MRinCMA*.

Step 4. Simulating *E*, *M* and *Y*.

Step 4.1: As [C4]

Step 4.2: As [C5]

Step 4.3:

$$Score = Y^*_{sim} = \hat{c}_0 + \hat{c}_1 E_{sim} + \hat{c}_2 M_{sim} + \hat{c}_3 Sex + \hat{c}_4 Age + \hat{c}_5 Smoke[1] + \hat{c}_6 Smoke[2] + \hat{c}_7 Smoke[3] + \varepsilon_{Y_{sim}}$$

where E_{sim} and M_{sim} are the values obtained from [C4] and [C5], respectively.

We recovered Y inverting the corresponding transformation.

$$Y_{sim} \sim Binomial\left(1, g^{-1}\left(Y^{*}{}_{sim}\right)\right) \quad [C7]$$

Denoting g the cumulative distribution function of the standard normal distribution or the logistic distribution, as appropiated.

S3: *Y* continuous, *E* continuous and *M* binary

Step 1. Selecting effect sizes based on the real dataset, PanGenEU.

To select the effect sizes, we fitted the following regression models for E, body mass index, M, long-standing diabetes mellitus and Y, pancreatic cancer, respectively:

$$E^{*} = a_0 + a_1 Sex + a_2 Smoke[1] + a_3 Smoke[2] + a_4 Smoke[3] + a_5 Age + a_6 IV_1 + \varepsilon_E \quad [C8]$$

$$M^{*} = b_0 + b_1 Sex + b_2 Smoke[1] + b_3 Smoke[2] + b_4 Smoke[3] + b_5 Age + b_6 IV_2 + b_7 E + \varepsilon_M \quad [C9]$$

$$Y^{*} = c_0 + c_1 E + c_2 M + c_3 Sex + c_4 Age + c_5 Smoke[1] + c_6 Smoke[2] + c_7 Smoke[3] + \varepsilon_Y \quad [C10]$$

M^{*} and Y^{*} corresponds to the probit score of long-standing diabetes mellitus and pancreatic cancer, respectively.

Step 4. Simulating *E*, *M* and *Y*.

Step 4.1: As [C4]

Step 4.2:

$$Score = M_{sim}{}^{*} = \hat{b}_0 + \hat{b}_1 Sex + \hat{b}_2 Smoke[1] + \hat{b}_3 Smoke[2] + \hat{b}_4 Smoke[3] + \hat{b}_5 Age + \hat{b}_6 IV_2 + \hat{b}_7 E_{sim} + \varepsilon_{M_{sim}}$$

where E_{sim} are the values obtained from [C4].

We recovered M inverting the corresponding transformation.

$$M_{sim} \sim Binomial\left(1, g^{-1}\left(M^{*}{}_{sim}\right)\right) \quad [C11]$$

Denoting g the cumulative distribution function of the standard normal distribution.

Step 4.3: As [C6]

S4: *Y* binary, *E* continuous and *M* binary

Step 1. Selecting effect sizes based on the real dataset, PanGenEU.

As [C7], [C8] and [C9], but in this case M^* and Y^* correspond to the probit score in SEM and logit score in *MRinCMA*.

Step 4.

Step 4.1: As [C4]

Step 4.2: Similar to [C11] denoting g the cumulative distribution function of the standard normal distribution or the logistic distribution, as appropriated.

Step 4.3: Similar to [C7] denoting g the cumulative distribution function of the standard normal distribution or the logistic distribution, as appropriated.

S5: *Y* continuous, *E* binary and *M* binary

Step 1. Selecting effect sizes based on the real dataset, PanGenEU.

To select the effect sizes, we fitted the following regression models for E, obesity, M, long-standing diabetes mellitus and Y, pancreatic cancer, respectively:

$$E^* = a_0 + a_1 Sex + a_2 Smoke[1] + a_3 Smoke[2] + a_4 Smoke[3] + a_5 Age + a_6 IV_1 + \varepsilon_E \quad \text{[C12]}$$

$$M^* = b_0 + b_1 Sex + b_2 Smoke[1] + b_3 Smoke[2] + b_4 Smoke[3] + b_5 Age + b_6 IV_2 + b_7 E + \varepsilon_M \quad \text{[C13]}$$

$$Y^* = c_0 + c_1 E + c_2 M + c_3 Sex + c_4 Age + c_5 Smoke[1] + c_6 Smoke[2] + c_7 Smoke[3] + \varepsilon_Y \quad \text{[C14]}$$

E^*, M^* and Y^* corresponds to the probit score of obsesity, long-standing diabetes mellitus and pancreatic cancer, respectively.

Step 4. Simulating *E*, *M* and *Y*.

Step 4.1.1:

$$Score = E_{sim}{}^* = \hat{a}_0 + \hat{a}_1 Sex + \hat{a}_2 Smoke[1] + \hat{a}_3 Smoke[2] + \hat{a}_4 Smoke[3] + \hat{a}_5 Age + \hat{a}_6 IV_1 + \varepsilon_{Esim} \quad \text{[C15]}$$

We recovered *M* inverting the corresponding transformation.

$$E_{sim} \sim Binomial\left(1, g^{-1}\left(E^*_{sim}\right)\right) \quad \text{[C16]}$$

Denoting g the cumulative distribution function of the standard normal distribution.

Step 4.1.2: As [C11]

Step 4.1.3: As [C7]

S6: *Y* binary, *E* binary and *M* binary

Step 1. Selecting effect sizes based on the real dataset, PanGenEU.

As [C12], [C13] and [C14], but in this case E^*, M^* and Y^* correspond to the probit score in SEM and logit score in *MRinCMA*.

Step 4. Simulating E, M and Y.

Step 4.1 As [C15] denoting g the cumulative distribution function of the standard normal distribution or the logistic distribution, as appropiated.

Step 4.2 As [C11] denoting g the cumulative distribution function of the standard normal distribution or the logistic distribution, as appropiated.

Step 4.3. As [C7] denoting g the cumulative distribution function of the standard normal distribution or the logistic distribution, as appropiated.

For the six scenarios (S1-S6) we simulated datasets of different sample sizes (i.e., $1{,}000$, $2{,}500$ and $10{,}000$) and we repeated the same simulation process across $\mathrm{m} = 2{,}000$ replications.

Appendix D. Additional Tables Chapter 5

Table D1. True coefficients considered in each scenario by method based on the effect sizes of PanGenEU

Based on PanGenEU	*MRinCMA*			SEM		
	Direct	Indirect	Total	Direct	Indirect	Total
S1	0.024	0.045	0.070	0.024	0.045	0.070
S2	0.042[a]	0.134	0.006	0.024[b]	0.045	0.001
S3	0.024[b]	0.045	0.070	0.024[b]	0.045	0.070
S4	0.042[a]	0.134	0.006	0.024[b]	0.045	0.001
S5	0.129[b]	0.133	0.261	0.129[b]	0.133	0.261
S6	0.199[a]	0.774	0.154	0.129[b]	0.133	0.017

MRinCMA= Mendelian Randomization in Causal Mediation Analysis. SEM = Structural Equation Models. S1: Y, E and M continuous; S2: Y binary and E and M continuous; S3: Y continuous, E continuous and M binary; S4: Y binary, E continuous and M binary; S5: Y continuous and E and M binary; S6: Y, E and M binary. [a]Logit coefficients; [b]Probit coefficients

Table D2. 95% CI coverage and bias-eliminated coverage for estimation of direct, indirect, and total effects for S1, S2, S3, S4, S5 and S6, considering a sample size of n=10,000

n = 10,000													
S1	**Direct**		**Indirect**		**Total**		**S2**	**Direct**		**Indirect**		**Total**	
	C	**BE-C**	**C**	**BE-C**	**C**	**BE-C**		**C**	**BE-C**	**C**	**BE-C**	**C**	**BE-C**
4SLS	99.40%	99.40%	93.70%	93.80%	92.00%	92.10%	***4SLS***	99.00%	99.00%	93.00%	93.00%	92.00%	92.00%
3SLS	99.40%	99.40%	92.20%	93.60%	91.50%	91.75%	***3SLS***	99.00%	99.00%	92.00%	93.00%	92.00%	92.00%
4SRI	88.30%	88.65%	93.30%	92.70%	92.60%	92.30%	***4SRI***	88.00%	89.00%	94.00%	94.00%	92.00%	92.00%
3SRI	88.30%	88.65%	91.40%	92.65%	92.50%	92.50%	***3SRI***	88.00%	89.00%	92.00%	93.00%	92.00%	92.00%
SEM	94.70%	94.78%	91.40%	92.00%	92.10%	91.95%	***SEM***	96.00%	97.00%	78.00%	94.00%	96.00%	99.00%
S3	**Direct**		**Indirect**		**Total**		**S4**	**Direct**		**Indirect**		**Total**	
	C	**BE-C**	**C**	**BE-C**	**C**	**BE-C**		**C**	**BE-C**	**C**	**BE-C**	**C**	**BE-C**
4SLS	96.70%	96.70%	88.20%	94.00%	88.90%	93.60%	**4SLS**	93.00%	93.00%	90.00%	93.00%	59.00%	92.00%
3SLS	96.70%	96.70%	83.50%	93.00%	85.60%	93.40%	**3SLS**	93.00%	93.00%	86.00%	92.00%	60.00%	93.00%
4SRI	89.40%	91.20%	92.60%	94.10%	93.80%	93.60%	**4SRI**	93.00%	93.00%	94.00%	94.00%	28.00%	93.00%
3SRI	89.40%	91.20%	93.40%	93.20%	92.60%	93.10%	**3SRI**	93.00%	93.00%	92.00%	94.00%	32.00%	93.00%
SEM	95.10%	95.10%	1.10%	93.00%	36.10%	93.30%	**SEM**	94.00%	96.00%	24.00%	94.00%	73.00%	99.00%
S5	**Direct**		**Indirect**		**Total**		**S6**	**Direct**		**Indirect**		**Total**	
	C	**BE-C**	**C**	**BE-C**	**C**	**BE-C**		**C**	**BE-C**	**C**	**BE-C**	**C**	**BE-C**
4SLS	95.50%	95.60%	97.10%	97.40%	93.60%	93.70%	***4SLS***	94.00%	94.00%	91.00%	93.00%	86.00%	94.00%
3SLS	95.50%	95.60%	96.00%	98.30%	93.30%	93.40%	***3SLS***	94.00%	94.00%	86.00%	93.00%	89.00%	94.00%
4SRI	89.10%	93.20%	97.00%	99.40%	84.10%	94.60%	***4SRI***	93.00%	93.00%	92.00%	95.00%	92.00%	94.00%
3SRI	89.10%	93.20%	97.20%	99.60%	83.60%	94.70%	***3SRI***	93.00%	93.00%	88.00%	97.00%	92.00%	94.00%
SEM	95.70%	96.40%	1.60%	92.50%	45.00%	92.40%	***SEM***	89.00%	96.00%	0.00%	98.00%	42.00%	99.00%

C = 95% CI coverage rate, BE-C: bias-eliminated coverage;4SLS=4-Stage-Least-Squares; 3SLS=3-Stage-Least-Squares; 4SRI=4-Stage-Residual-Inclusion; 3SRI=3-Stage-Residual-Inclusion; SEM= Structural Equation Models.

Table D3. Bias and MSE for S3, S4, S5 and S6, considering a sample size of n=10,000

n = 10,000	Direct			Indirect			Total		
S3	**Bias**	**MSE**	**EmpSE**	**Bias**	**MSE**	**EmpSE**	**Bias**	**MSE**	**EmpSE**
	Estimate±SE	**Estimate±SE**	**Estimate±SE**	**Estimate±SE**	**Estimate±SE**	**Estimate±SE**	**Estimate±SE**	**Estimate±SE**	**Estimate±SE**
***4SLS*[b]**	0.000±0.000	0.000±0.000	0.016±0.000	-0.011±0.000	0.000±0.000	0.019±0.000	-0.011±0.000	0.001±0.000	0.023±0.000
***3SLS*[b]**	0.000±0.000	0.000±0.000	0.016±0.000	-0.014±0.000	0.000±0.000	0.017±0.000	0.015±0.000	0.001±0.000	0.022±0.000
***4SRI*[b]**	-0.004±0.000	0.000±0.000	0.015±0.000	0.006±0.000	0.000±0.000	0.015±0.000	0.001 ±0.000	0.001±0.000	0.023±0.000
***3SRI*[b]**	-0.004±0.000	0.000±0.000	0.015±0.000	0.000±0.000	0.001±0.000	0.0151±0.000	-0.004±0.000	0.001±0.000	0.023±0.000
***SEM*[b]**	0.000±0.000	0.000±0.000	0.016±0.000	-0.035±0.000	0.002±0.000	0.006±0.000	-0.036±0.000	0.002±0.000	0.017±0.000
S4	**Bias**	**MSE**	**EmpSE**	**Bias**	**MSE**	**EmpSE**	**Bias**	**MSE**	**EmpSE**
	Estimate±SE	**Estimate±SE**	**Estimate±SE**	**Estimate±SE**	**Estimate±SE**	**Estimate±SE**	**Estimate±SE**	**Estimate±SE**	**Estimate±SE**
4SLS[a]	-0.003±0.001	0.002±0.000	0.048±0.001	-0.034±0.003	0.008±0.000	0.086±0.001	0.132±0.003	0.024±0.001	0.088±0.001
3SLS[a]	-0.002±0.001	0.002±0.000	0.048±0.001	-0.045±0.002	0.008±0.000	0.078±0.001	0.121±0.002	0.020±0.001	0.080±0.001
4SRI[a]	-0.002±0.001	0.002±0.000	0.047±0.001	0.002±0.002	0.004±0.000	0.061±0.001	0.169±0.002	0.034±0.001	0.051±0.001
3SRI[a]	-0.002±0.001	0.002±0.000	0.047±0.001	-0.011±0.002	0.003±0.000	0.057±0.001	0.155±0.002	0.029±0.001	0.046±0.001
SEM[b]	0.008±0.001	0.001±0.000	0.027±0.001	0.008±0.001	0.001±0.000	0.008±0.001	-0.001±0.000	0.000±0.000	0.009±0.001
S5	**Bias**	**MSE**	**EmpSE**	**Bias**	**MSE**	**EmpSE**	**Bias**	**MSE**	**EmpSE**
	Estimate±SE	**Estimate±SE**	**Estimate±SE**	**Estimate±SE**	**Estimate±SE**	**Estimate±SE**	**Estimate±SE**	**Estimate±SE**	**Estimate±SE**
4SLS[b]	-0.009 ±0.006	0.075±0.002	0.274±0.004	-0.002±0.004	0.026±0.001	0.161±0.002	-0.011±0.007	0.102±0.003	0.320±0.005
3SLS[b]	-0.009 ±0.006	0.075±0.002	0.274±0.004	-0.034±0.003	0.023±0.001	0.145±0.002	-0.043±0.007	0.101±0.003	0.316±0.005
4SRI[b]	-0.211±0.008	0.162±0.005	0.342±0.005	-0.115±0.003	0.035±0.001	0.148±0.002	-0.326±0.008	0.241±0.006	0.366±0.006
3SRI[b]	-0.211±0.008	0.162±0.005	0.342±0.005	-0.124±0.003	0.003±0.001	0.129±0.002	-0.335±0.008	0.241±0.007	0.359±0.006
SEM[b]	-0.009±0.001	0.004±0.000	0.062±0.001	-0.117±0.000	0.014±0.000	0.021±0.000	-0.127±0.001	0.020±0.000	0.066±0.001
S6	**Bias**	**MSE**	**EmpSE**	**Bias**	**MSE**	**EmpSE**	**Bias**	**MSE**	**EmpSE**
	Estimate±SE	**Estimate±SE**	**Estimate±SE**	**Estimate±SE**	**Estimate±SE**	**Estimate±SE**	**Estimate±SE**	**Estimate±SE**	**Estimate±SE**
4SLS[a]	-0.057±0.018	0.327±0.016	0.570±0.001	-0.134±0.018	0.330±0.015	0.559±0.001	0.628±0.024	0.957±0.040	0.489±0.001
3SLS[a]	-0.057±0.018	0.327±0.016	0.570±0.001	-0.254±0.016	0.316±0.012	0.501±0.001	0.508±0.023	0.771±0.033	0.439±0.001
4SRI[a]	-0.045±0.023	0.521±0.025	0.721±0.001	-0.129±0.023	0.569±0.033	0.744±0.001	0.645±0.032	1.429±0.069	0.717±0.001
3SRI[a]	-0.045±0.023	0.521±0.025	0.721±0.001	-0.301±0.021	0.527±0.026	0.661±0.001	0.473±0.030	1.134±0.056	0.682±0.001
SEM[b]	-0.107±0.004	0.026±0.001	0.270±0.001	-0.126±0.000	0.016±0.000	0.125±0.001	-0.017±0.000	0.000±0.000	0.027±0.001

MSE=Mean Squared Error; EmpSE=Empirical Standard Error; SE= standard error;4SLS=4-Stage-Least-Squares; 3SLS=3-Stage-Least-Squares; 4SRI=4-Stage-Residual-Inclusion; 3SRI=3-Stage-Residual-Inclusion; SEM= Structural Equation Models. [a]Logit model were used. [b]Probit model were used, when appropriated.

Appendix E. Performance measures formulas

1. Bias:

$$\text{Bias} = \frac{1}{n_{sim}}\sum_{i=1}^{n_{sim}}\hat{\theta}_i - \theta \quad \text{[E1]}$$

where n_{sim} is the total number of simulations (runs), $\hat{\theta}_i$ are the parameter estimates and θ is the real estimand. The standard error can be obtained with the following formula:

$$\text{Bias}_{\text{se}} = \sqrt{\frac{1}{n_{sim}\left(n_{sim}-1\right)}\sum_{i=1}^{n_{sim}}\left(\hat{\theta}_i - \bar{\theta}\right)^2} \quad \text{[E2]}$$

2. Mean Squared Error (MSE)

$$\text{MSE} = \frac{1}{n_{sim}}\sum_{i=1}^{n_{sim}}\left(\hat{\theta}_i - \theta\right) \quad \text{[E3]}$$

and

$$\text{MSE}_{\text{SE}} = \sqrt{\frac{\sum_{i=1}^{n_{sim}}\left[\left(\hat{\theta}_i - \theta\right)^2 - \hat{\text{MSE}}\right]^2}{n_{sim}\left(n_{sim}-1\right)}} \quad \text{[E4]}$$

3. Coverage and Bias-eliminated Coverage

$$\text{coverage} = \frac{1}{n_{sim}}\sum_{i=1}^{n_{sim}}1\left(\hat{\theta}_{low,i} \le \theta \le \hat{\theta}_{upp,i}\right) \quad \text{[E5]}$$

$$\text{bias\_eliminated\_coverage} = \frac{1}{n_{sim}}\sum_{i=1}^{n_{sim}}1\left(\hat{\theta}_{low,i} \le \bar{\theta} \le \hat{\theta}_{upp,i}\right) \quad \text{[E6]}$$

4. Empirical Standard Error (EmpSE)

$$\text{EmpSE} = \sqrt{\frac{1}{\left(n_{sim}-1\right)}\sum_{i=1}^{n_{sim}}\left(\hat{\theta}_i - \bar{\theta}\right)^2} \quad \text{[E7]}$$

$$\text{EmpSE}_{\text{error}} = \frac{\hat{\text{EmpSE}}}{\sqrt{2\left(n_{sim}-1\right)}} \quad \text{[E8]}$$

Bias (B), can be obtained as the men difference of estimates from the real parameter value, where n is the total number of simulations (runs), $\hat{\theta}_i$ are the simulated estimates and θ is the real estimand. A value close to zero represents less bias, indicating that estimates obtained in the simulations are similar to the estimand that we are setting.

The mean squared error (MSE) is the sum of the squared bias and variance of $\hat{\theta}$. As in the case of bias, we aim to obtain a value as close to zero as possible, indicating a good performance.

The empirical standard error (EmpSE) indicates the precision of the estimator to θ. The EmpSE estimates the standard deviation of the $\hat{\theta}_i$ across the n_{sim} runs.

On the other hand, coverage measures the probability that the confidence intervals from all estimations obtained in the n simulations contains the real value of the estimand.

Bias-eliminated coverage measures the probability that the confidence interval of all simulated estimands contains the mean value of the same estimands.

These measures were calculated for direct, indirect, and total effects, respectively.

There is a close connection between these measures. When bias is high, it is possible that the coverage results do not reach the nominal levels. It is important to consider also the bias-eliminated coverage, being an extension of coverage that should be used when bias is present.

Appendix F. Additional Tables Chapter 6

Table F1. Scenario A1, results for positive, negative, and mixed confounding, adjustment approach, α1=0.3

		Positive confounding (α_2 and β_2 = 0.8)		Negative confounding (α_2 and β_2 = -0.8)		Mixed confounding (α_2 =0.8 and β_2 = -0.8)	
		Median of causal estimates	Type I error rate (%)	Median of causal estimates	Type I error rate (%)	Median of causal estimates	Type I error rate (%)
μ_1	μ_2	*No adjust for collider*					
-0.5	-0.5	0.00	6%	0.00	5%	0.00	6%
	0	0.01	4%	0.00	6%	0.00	2%
	0.5	0.00	5%	0.00	4%	0.00	5%
0	-0.5	0.00	4%	0.00	4%	0.00	4%
	0	0.00	5%	0.00	5%	0.00	6%
	0.5	0.00	4%	0.00	5%	0.00	5%
0.5	-0.5	0.00	4%	0.00	5%	0.00	5%
	0	0.00	7%	0.00	4%	0.00	4%
	0.5	0.00	6%	0.00	6%	0.00	4%
μ_1	μ_2	*Adjust Y/G for collider*					
-0.5	-0.5	-0.17	100%	0.04	13%	0.18	100%
	0	-0.12	81%	-0.12	77%	0.11	76%
	0.5	0.03	11%	-0.18	100%	-0.03	10%
0	-0.5	0.00	6%	0.00	5%	0.00	6%
	0	0.00	4%	0.00	5%	0.00	5%
	0.5	0.00	4%	0.00	4%	0.00	6%
0.5	-0.5	0.04	13%	-0.18	99%	-0.03	12%
	0	-0.11	75%	-0.11	76%	0.11	77%
	0.5	-0.17	99%	0.03	9%	0.18	100%

- Empirical Type I error rate represents the proportion of simulated datasets where the null hypothesis is not rejected

Table F2. Scenario A1, results for positive, negative, and mixed confounding, adjustment approach, α1=0.05

		Positive confounding (α_2 and β_2 = 0.8)		Negative confounding (α_2 and β_2 = -0.8)		Mixed confounding (α_2 =0.8 and β_2 = -0.8)	
		Median of causal estimates	Type I error rate (%)	Median of causal estimates	Type I error rate (%)	Median of causal estimates	Type I error rate (%)
μ_1	μ_2	*No adjust for collider*					
-0.5	-0.5	-0.02	3%	0.00	7%	-0.03	5%
	0	-0.01	5%	0.01	5%	-0.01	4%
	0.5	-0.01	5%	0.00	5%	0.03	5%
0	-0.5	0.05	5%	0.00	3%	0.00	6%
	0	-0.01	4%	-0.01	5%	-0.02	5%
	0.5	0.02	5%	0.00	5%	-0.04	6%
0.5	-0.5	0.00	7%	0.00	5%	0.00	5%
	0	-0.01	4%	-0.01	5%	-0.01	4%
	0.5	-0.01	5%	0.03	3%	0.00	4%
μ_1	μ_2	*Adjust Y/G for collider*					
-0.5	-0.5	-0.18	11%	0.04	5%	0.16	13%
	0	-0.13	8%	-0.11	6%	0.13	7%
	0.5	0.02	5%	-0.18	13%	-0.05	6%
0	-0.5	-0.01	4%	-0.02	6%	-0.02	5%
	0	0.02	5%	0.00	5%	-0.04	6%
	0.5	0.00	6%	0.01	7%	-0.01	5%
0.5	-0.5	0.02	5%	-0.19	13%	-0.04	4%
	0	-0.12	7%	-0.08	6%	0.11	8%
	0.5	-0.16	9%	0.03	5%	0.17	12%

Empirical Type I error rate represents the proportion of simulated datasets where the null hypothesis is not rejected

Table F3. Scenario A2, results for negative, and mixed confounding, stratification approach, α1=0.1

		Negative confounding (α_2 and β_2 = -0.8)									
		Stratifying on collider, C					Stratifying on residual collider, C_0				
μ_1	μ_2	Proportion homogeneity rejected (%)	Median estimates Q1	Median estimates Q2	Median estimates Q3	Median estimates Q4	Proportion homogeneity rejected (%)	Median estimates Q1	Median estimates Q2	Median estimates Q3	Median estimates Q4
-0.5	-0.5	6%	0.44	0.45	0.46	0.45	3%	0.48	0.48	0.51	0.52
	0	4%	0.28	0.25	0.24	0.32	5%	0.49	0.49	0.49	0.5
	0.5	6%	0.26	0.17	0.19	0.22	6%	0.52	0.49	0.5	0.46
0	-0.5	4%	0.51	0.51	0.50	0.49	4%	0.52	0.51	0.49	0.49
	0	7%	0.49	0.50	0.50	0.49	5%	0.5	0.49	0.5	0.49
	0.5	4%	0.47	0.50	0.49	0.48	5%	0.47	0.49	0.5	0.48
0.5	-0.5	7%	0.23	0.18	0.18	0.24	7%	0.47	0.48	0.51	0.49
	0	4%	0.28	0.27	0.25	0.32	5%	0.47	0.5	0.5	0.51
	0.5	6%	0.46	0.44	0.46	0.47	6%	0.48	0.5	0.49	0.53
		Mixed confounding (α_2 =0.8 and β_2 = -0.8)									
-0.5	-0.5	6%	0.54	0.53	0.52	0.52	5%	0.52	0.51	0.49	0.49
	0	6%	0.47	0.45	0.47	0.47	5%	0.51	0.48	0.48	0.50
	0.5	5%	0.43	0.39	0.38	0.39	5%	0.51	0.51	0.48	0.49
0	-0.5	7%	0.47	0.50	0.50	0.50	7%	0.47	0.50	0.50	0.49
	0	5%	0.50	0.49	0.47	0.49	4%	0.49	0.49	0.48	0.48
	0.5	5%	0.50	0.52	0.51	0.50	5%	0.50	0.51	0.51	0.49
0.5	-0.5	5%	0.42	0.39	0.41	0.43	6%	0.51	0.51	0.50	0.54
	0	5%	0.46	0.46	0.48	0.47	5%	0.49	0.49	0.51	0.50
	0.5	5%	0.51	0.52	0.54	0.54	4%	0.48	0.50	0.51	0.51

Proportion homogeneity rejected represents the proportion of simulated datasets where the null hypothesis of homogeneity is rejected

Table F4. Scenario A2, results for positive, negative, and mixed confounding, stratification approach, α1=0.3

		Positive confounding (α_2 and β_2 = 0.8)									
		Stratifying on collider, C					Stratifying on residual collider, C_0				
μ_1	μ_2	Proportion homogeneity rejected (%)	Median estimates Q1	Median estimates Q2	Median estimates Q3	Median estimates Q4	Proportion homogeneity rejected (%)	Median estimates Q1	Median estimates Q2	Median estimates Q3	Median estimates Q4
-0.5	-0.5	11%	0.25	0.19	0.19	0.26	5%	0.50	0.49	0.50	0.50
	0	7%	0.31	0.25	0.25	0.30	7%	0.51	0.50	0.49	0.50
	0.5	5%	0.47	0.44	0.46	0.46	5%	0.50	0.50	0.50	0.50
0	-0.5	4%	0.51	0.50	0.50	0.50	5%	0.51	0.50	0.50	0.50
	0	5%	0.50	0.51	0.50	0.50	5%	0.50	0.51	0.50	0.50
	0.5	6%	0.51	0.49	0.50	0.50	6%	0.50	0.49	0.50	0.50
0.5	-0.5	6%	0.47	0.46	0.45	0.46	5%	0.51	0.51	0.50	0.50
	0	8%	0.30	0.25	0.25	0.30	6%	0.50	0.51	0.50	0.49
	0.5	6%	0.26	0.20	0.18	0.25	4%	0.50	0.50	0.49	0.50
		Negative confounding (α_2 and β_2 = -0.8)									
-0.5	-0.5	6%	0.46	0.45	0.46	0.46	5%	0.49	0.50	0.50	0.50
	0	7%	0.31	0.24	0.25	0.31	5%	0.50	0.50	0.50	0.50
	0.5	8%	0.25	0.19	0.18	0.25	6%	0.49	0.50	0.50	0.50
0	-0.5	3%	0.50	0.51	0.49	0.49	4%	0.50	0.50	0.49	0.50
	0	4%	0.50	0.51	0.49	0.50	4%	0.50	0.50	0.49	0.50
	0.5	4%	0.50	0.51	0.51	0.49	3%	0.50	0.51	0.51	0.49
0.5	-0.5	8%	0.26	0.19	0.19	0.24	6%	0.50	0.50	0.51	0.49
	0	8%	0.31	0.25	0.24	0.31	4%	0.50	0.51	0.49	0.49
	0.5	3%	0.46	0.45	0.45	0.46	5%	0.50	0.50	0.49	0.50
		Mixed confounding (α_2 =0.8 and β_2 = -0.8)									
-0.5	-0.5	4%	0.52	0.53	0.53	0.52	4%	0.50	0.50	0.50	0.50
	0	6%	0.48	0.48	0.46	0.47	4%	0.51	0.51	0.51	0.49
	0.5	4%	0.41	0.39	0.39	0.42	3%	0.50	0.50	0.50	0.50
0	-0.5	7%	0.50	0.50	0.51	0.50	7%	0.50	0.51	0.51	0.50
	0	5%	0.50	0.50	0.50	0.50	5%	0.50	0.50	0.50	0.50
	0.5	5%	0.50	0.50	0.49	0.50	5%	0.49	0.50	0.49	0.50
0.5	-0.5	5%	0.41	0.39	0.39	0.41	5%	0.50	0.50	0.51	0.50
	0	4%	0.48	0.47	0.47	0.48	4%	0.50	0.50	0.50	0.50
	0.5	6%	0.52	0.54	0.52	0.52	6%	0.49	0.51	0.50	0.50

Proportion homogeneity rejected represents the proportion of simulated datasets where the null hypothesis of homogeneity is rejected

Table F5. Scenario A2, results for positive, negative, and mixed confounding, stratification approach, α1=0.05

		Positive confounding (α_2 and β_2 = 0.8)									
		Stratifying on collider, C					Stratifying on residual collider, C_0				
μ_1	μ_2	Proportion homogeneity rejected (%)	Median estimates Q1	Median estimates Q2	Median estimates Q3	Median estimates Q4	Proportion homogeneity rejected (%)	Median estimates Q1	Median estimates Q2	Median estimates Q3	Median estimates Q4
-0.5	-0.5	7%	0.29	0.21	0.19	0.25	6%	0.53	0.51	0.48	0.50
	0	6%	0.24	0.26	0.23	0.32	5%	0.45	0.50	0.48	0.51
	0.5	6%	0.37	0.48	0.39	0.41	5%	0.45	0.56	0.46	0.47
0	-0.5	6%	0.48	0.48	0.53	0.51	6%	0.49	0.51	0.49	0.52
	0	4%	0.49	0.47	0.54	0.47	4%	0.49	0.49	0.55	0.47
	0.5	5%	0.57	0.49	0.53	0.49	5%	0.55	0.51	0.58	0.48
0.5	-0.5	8%	0.44	0.50	0.45	0.48	8%	0.50	0.53	0.47	0.54
	0	6%	0.31	0.20	0.27	0.31	6%	0.52	0.44	0.54	0.53
	0.5	6%	0.24	0.16	0.16	0.23	5%	0.49	0.51	0.50	0.48
		Negative confounding (α_2 and β_2 = -0.8)									
-0.5	-0.5	5%	0.49	0.43	0.45	0.48	5%	0.50	0.45	0.47	0.48
	0	5%	0.28	0.21	0.25	0.28	5%	0.46	0.50	0.53	0.47
	0.5	4%	0.30	0.22	0.14	0.29	4%	0.54	0.52	0.49	0.56
0	-0.5	4%	0.49	0.52	0.48	0.52	5%	0.49	0.55	0.47	0.54
	0	4%	0.55	0.46	0.46	0.48	4%	0.53	0.48	0.49	0.46
	0.5	6%	0.48	0.49	0.47	0.57	6%	0.47	0.50	0.49	0.58
0.5	-0.5	7%	0.26	0.22	0.19	0.24	6%	0.51	0.48	0.52	0.49
	0	4%	0.32	0.23	0.33	0.33	5%	0.52	0.51	0.59	0.52
	0.5	6%	0.49	0.40	0.40	0.46	6%	0.52	0.50	0.45	0.47
		Mixed confounding (α_2 = 0.8 and β_2 = -0.8)									
-0.5	-0.5	6%	0.51	0.52	0.55	0.49	5%	0.48	0.49	0.51	0.49
	0	6%	0.47	0.46	0.41	0.50	6%	0.50	0.51	0.46	0.52
	0.5	6%	0.44	0.39	0.35	0.38	6%	0.54	0.51	0.51	0.46
0	-0.5	6%	0.50	0.52	0.49	0.53	6%	0.52	0.50	0.50	0.53
	0	6%	0.50	0.55	0.53	0.49	5%	0.48	0.55	0.53	0.50
	0.5	6%	0.51	0.51	0.45	0.51	6%	0.53	0.50	0.47	0.54
0.5	-0.5	5%	0.45	0.41	0.32	0.42	6%	0.53	0.53	0.43	0.50
	0	5%	0.49	0.46	0.45	0.45	5%	0.51	0.52	0.46	0.46
	0.5	5%	0.49	0.54	0.50	0.54	4%	0.49	0.48	0.49	0.52

Proportion homogeneity rejected represents the proportion of simulated datasets where the null hypothesis of homogeneity is rejected

Table F6. Scenario A2, results for positive confounding, stratification approach, when α1=0.5 and n=50,000, respectively

		Positive confounding (α_2 and β_2 = 0.8), α_1=0.5, n =10,000									
		Stratifying on collider, C					Stratifying on residual collider, C_0				
μ_1	μ_2	Proportion homogeneity rejected (%)	Median estimates Q1	Median estimates Q2	Median estimates Q3	Median estimates Q4	Proportion homogeneity rejected (%)	Median estimates Q1	Median estimates Q2	Median estimates Q3	Median estimates Q4
-0.5	-0.5	16%	0.26	0.19	0.19	0.26	6%	0.5	0.5	0.5	0.5
	0	10%	0.31	0.25	0.25	0.31	5%	0.5	0.5	0.5	0.5
	0.5	5%	0.46	0.45	0.45	0.46	4%	0.5	0.5	0.5	0.5
0	-0.5	5%	0.5	0.5	0.5	0.5	5%	0.5	0.5	0.5	0.5
	0	6%	0.5	0.5	0.5	0.5	5%	0.49	0.5	0.5	0.5
	0.5	4%	0.5	0.51	0.5	0.5	6%	0.5	0.51	0.5	0.5
0.5	-0.5	4%	0.46	0.45	0.45	0.46	4%	0.5	0.5	0.5	0.5
	0	9%	0.31	0.25	0.25	0.31	6%	0.5	0.49	0.5	0.5
	0.5	14%	0.26	0.18	0.19	0.26	5%	0.5	0.5	0.5	0.5
		Positive confounding (α_2 and β_2 = 0.8), α_1=0.1, n =50,000									
-1	-1	11%	0.12	0	0.01	0.13	6%	0.51	0.49	0.49	0.53
	0	16%	0.09	-0.04	-0.02	0.09	5%	0.51	0.51	0.5	0.49
	1	4%	0.36	0.4	0.38	0.39	3%	0.49	0.48	0.53	0.5
-0.5	-1	9%	0.27	0.21	0.22	0.27	4%	0.47	0.51	0.51	0.5
	0	7%	0.31	0.23	0.26	0.32	5%	0.5	0.49	0.5	0.52
	1	4%	0.59	0.63	0.61	0.62	2%	0.48	0.48	0.47	0.49
0	-1	7%	0.5	0.49	0.49	0.5	10%	0.49	0.49	0.51	0.5
	0	6%	0.48	0.52	0.53	0.51	6%	0.47	0.52	0.52	0.51
	1	4%	0.48	0.53	0.5	0.5	3%	0.48	0.54	0.49	0.51
0.5	-1	4%	0.61	0.65	0.65	0.63	7%	0.5	0.48	0.52	0.5
	0	7%	0.31	0.29	0.23	0.29	4%	0.51	0.52	0.48	0.48
	1	6%	0.27	0.21	0.21	0.28	5%	0.48	0.51	0.5	0.49
1	-1	5%	0.4	0.37	0.37	0.39	7%	0.5	0.49	0.49	0.49
	0	10%	0.08	-0.04	-0.03	0.06	6%	0.49	0.49	0.49	0.5
	1	13%	0.12	0	0.01	0.11	4%	0.52	0.5	0.51	0.49

Proportion homogeneity rejected represents the proportion of simulated datasets where the null hypothesis of homogeneity is rejected

Table F7. Scenario A3, results for negative, and mixed confounding, stratification approach, α_1=0.1

		Negative confounding (α_2 and β_2 = -0.8)									
		Stratifying on collider, C					Stratifying on residual collider, C_0				
μ_1	μ_2	Proportion homogeneity rejected (%)	Median estimates Q1	Median estimates Q2	Median estimates Q3	Median estimates Q4	Proportion homogeneity rejected (%)	Median estimates Q1	Median estimates Q2	Median estimates Q3	Median estimates Q4
-0.5	-0.5	15%	0.22	0.45	0.55	0.76	21%	0.22	0.45	0.55	0.76
	0	19%	0.10	0.23	0.33	0.61	29%	0.10	0.23	0.33	0.61
	0.5	24%	-0.03	0.14	0.30	0.64	50%	-0.03	0.14	0.30	0.64
0	-0.5	17%	0.31	0.53	0.68	0.87	16%	0.31	0.53	0.68	0.87
	0	14%	0.34	0.54	0.61	0.88	13%	0.34	0.54	0.61	0.88
	0.5	15%	0.28	0.54	0.66	0.93	15%	0.28	0.54	0.66	0.93
0.5	-0.5	23%	0.10	0.23	0.36	0.72	42%	0.10	0.23	0.36	0.72
	0	13%	0.18	0.27	0.43	0.69	29%	0.18	0.27	0.43	0.69
	0.5	14%	0.32	0.47	0.66	0.89	22%	0.32	0.47	0.66	0.89
		Mixed confounding (α_2 = 0.8 and β_2 = -0.8)									
-0.5	-0.5	32%	0.22	0.45	0.55	0.76	63%	0.02	0.40	0.60	1.01
	0	18%	0.10	0.23	0.33	0.61	43%	0.14	0.40	0.62	0.89
	0.5	23%	-0.03	0.14	0.30	0.64	35%	0.15	0.42	0.61	0.84
0	-0.5	27%	0.31	0.53	0.68	0.87	27%	0.29	0.55	0.67	0.90
	0	23%	0.34	0.54	0.61	0.88	24%	0.35	0.54	0.67	0.89
	0.5	27%	0.28	0.54	0.66	0.93	27%	0.34	0.53	0.68	0.91
0.5	-0.5	35%	0.10	0.23	0.36	0.72	36%	0.38	0.62	0.79	1.03
	0	20%	0.18	0.27	0.43	0.69	43%	0.30	0.60	0.80	1.10
	0.5	26%	0.32	0.47	0.66	0.89	56%	0.25	0.56	0.81	1.17

Proportion homogeneity rejected represents the proportion of simulated datasets where the null hypothesis of homogeneity is rejected

Table F8. Scenario A3, results for positive, negative, and mixed confounding, stratification approach, α1=0.3

		Positive confounding (α_2 and β_2 = 0.8)									
		Stratifying on collider, C					Stratifying on residual collider, C_0				
μ_1	μ_2	Proportion homogeneity rejected (%)	Median estimates Q1	Median estimates Q2	Median estimates Q3	Median estimates Q4	Proportion homogeneity rejected (%)	Median estimates Q1	Median estimates Q2	Median estimates Q3	Median estimates Q4
-0.5	-0.5	98%	-0.04	0.15	0.29	0.61	100%	0.04	0.37	0.62	0.99
	0	88%	0.06	0.22	0.34	0.61	100%	0.10	0.40	0.61	0.88
	0.5	81%	0.24	0.42	0.55	0.76	96%	0.18	0.42	0.58	0.84
0	-0.5	88%	0.31	0.54	0.67	0.88	88%	0.31	0.54	0.67	0.88
	0	76%	0.35	0.55	0.65	0.85	76%	0.34	0.55	0.66	0.85
	0.5	90%	0.31	0.53	0.67	0.88	89%	0.31	0.53	0.68	0.88
0.5	-0.5	75%	0.34	0.52	0.64	0.87	93%	0.37	0.61	0.78	1.03
	0	82%	0.17	0.30	0.43	0.70	99%	0.31	0.60	0.80	1.10
	0.5	98%	0.07	0.23	0.37	0.70	100%	0.23	0.58	0.82	1.17
		Negative confounding (α_2 and β_2 = -0.8)									
-0.5	-0.5	85%	0.23	0.43	0.55	0.79	97%	0.16	0.42	0.59	0.84
	0	86%	0.06	0.23	0.33	0.60	100%	0.11	0.40	0.59	0.88
	0.5	98%	-0.04	0.15	0.29	0.60	100%	0.02	0.38	0.62	0.97
0	-0.5	88%	0.31	0.53	0.66	0.88	88%	0.31	0.53	0.66	0.88
	0	76%	0.35	0.53	0.67	0.85	77%	0.35	0.53	0.66	0.85
	0.5	89%	0.31	0.52	0.67	0.89	88%	0.31	0.53	0.67	0.89
0.5	-0.5	97%	0.06	0.24	0.36	0.70	100%	0.23	0.58	0.82	1.16
	0	83%	0.16	0.30	0.42	0.71	99%	0.31	0.59	0.80	1.09
	0.5	77%	0.34	0.51	0.65	0.88	93%	0.37	0.61	0.79	1.03
		Mixed confounding (α_2 = 0.8 and β_2 = -0.8)									
-0.5	-0.5	100%	0.23	0.50	0.63	0.87	100%	0.04	0.38	0.62	0.97
	0	99%	0.23	0.45	0.55	0.78	100%	0.10	0.40	0.60	0.89
	0.5	99%	0.18	0.37	0.50	0.72	100%	0.17	0.42	0.59	0.84
0	-0.5	100%	0.31	0.52	0.67	0.89	100%	0.31	0.52	0.66	0.89
	0	98%	0.35	0.53	0.66	0.86	98%	0.35	0.53	0.66	0.86
	0.5	99%	0.31	0.53	0.68	0.88	99%	0.31	0.53	0.68	0.88
0.5	-0.5	98%	0.30	0.46	0.58	0.83	100%	0.38	0.62	0.78	1.04
	0	98%	0.33	0.52	0.64	0.88	100%	0.31	0.60	0.80	1.09
	0.5	100%	0.32	0.56	0.71	0.97	100%	0.23	0.58	0.82	1.17

Proportion homogeneity rejected represents the proportion of simulated datasets where the null hypothesis of homogeneity is rejected

Table F9. Scenario A3, results for positive, negative, and mixed confounding, stratification approach, α1=0.05

		Positive confounding (α_2 and β_2 = 0.8)									
		Stratifying on collider, C					Stratifying on residual collider, C_0				
μ_1	μ_2	Proportion homogeneity rejected (%)	Median estimates Q1	Median estimates Q2	Median estimates Q3	Median estimates Q4	Proportion homogeneity rejected (%)	Median estimates Q1	Median estimates Q2	Median estimates Q3	Median estimates Q4
-0.5	-0.5	10%	-0.02	0.14	0.32	0.63	17%	0.05	0.35	0.66	0.99
	0	8%	0.11	0.22	0.31	0.53	9%	0.14	0.38	0.60	0.85
	0.5	8%	0.22	0.44	0.58	0.77	9%	0.17	0.41	0.61	0.86
0	-0.5	8%	0.33	0.57	0.66	0.79	8%	0.33	0.55	0.66	0.81
	0	6%	0.27	0.48	0.65	0.84	6%	0.27	0.51	0.64	0.86
	0.5	7%	0.32	0.50	0.73	0.93	8%	0.33	0.51	0.75	0.93
0.5	-0.5	4%	0.37	0.48	0.66	0.89	7%	0.41	0.61	0.85	1.03
	0	6%	0.15	0.34	0.41	0.68	9%	0.27	0.60	0.79	1.05
	0.5	8%	0.03	0.26	0.41	0.71	16%	0.18	0.60	0.84	1.19
		Negative confounding (α_2 and β_2 = -0.8)									
-0.5	-0.5	8%	0.20	0.39	0.57	0.78	11%	0.14	0.38	0.61	0.88
	0	7%	0.10	0.17	0.37	0.56	10%	0.13	0.33	0.61	0.85
	0.5	7%	0.01	0.14	0.27	0.60	14%	0.07	0.37	0.61	0.95
0	-0.5	6%	0.28	0.55	0.67	0.89	8%	0.26	0.56	0.69	0.88
	0	7%	0.33	0.55	0.66	0.87	6%	0.35	0.55	0.65	0.90
	0.5	5%	0.27	0.56	0.68	0.87	5%	0.26	0.51	0.70	0.87
0.5	-0.5	8%	0.08	0.22	0.37	0.74	14%	0.26	0.55	0.80	1.19
	0	8%	0.16	0.29	0.45	0.67	10%	0.31	0.61	0.81	1.07
	0.5	6%	0.37	0.55	0.69	0.91	5%	0.40	0.66	0.82	1.08
		Mixed confounding (α_2 = 0.8 and β_2= -0.8)									
-0.5	-0.5	10%	0.25	0.50	0.64	0.87	20%	0.07	0.36	0.62	0.95
	0	7%	0.30	0.46	0.53	0.78	10%	0.16	0.42	0.58	0.91
	0.5	8%	0.18	0.35	0.53	0.73	13%	0.15	0.43	0.63	0.84
0	-0.5	7%	0.35	0.54	0.68	0.86	8%	0.35	0.54	0.67	0.88
	0	7%	0.33	0.52	0.63	0.89	8%	0.34	0.53	0.63	0.90
	0.5	9%	0.31	0.57	0.66	0.88	9%	0.30	0.58	0.65	0.88
0.5	-0.5	8%	0.29	0.42	0.54	0.77	9%	0.39	0.58	0.74	0.99
	0	9%	0.31	0.47	0.61	0.87	14%	0.26	0.55	0.78	1.09
	0.5	11%	0.31	0.55	0.75	0.99	20%	0.20	0.55	0.85	1.19

Proportion homogeneity rejected represents the proportion of simulated datasets where the null hypothesis of homogeneity is rejected

Table F10. Scenario B1, results for positive confounding, adjustment approach, α1=0.1, considering C as a function of both risk factor and outcome

		Positive confounding (α_2 and β_2= 0.8)			
		Median estimate	**Type I error rate (%)**	**Median estimate**	**Type I error rate (%)**
μ_1	**μ_3**	*No adjust for collider*		*Adjust Y/G for collider*	
-0.5	-0.5	-0.01	4%	-0.27	65%
	0	0.00	7%	-0.03	7%
	0.5	-0.01	5%	0.23	52%
0	-0.5	0.01	6%	0.00	5%
	0	0.00	6%	0.00	6%
	0.5	0.00	5%	0.01	7%
0.5	-0.5	-0.01	6%	0.08	8%
	0	-0.01	3%	-0.17	28%
	0.5	0.02	5%	-0.25	74%

Empirical Type I error rate represents the proportion of simulated datasets where the null hypothesis is not rejected

Table F11. Scenario B2 and B3, results for positive confounding, stratification approach, α1=0.1, considering C as a function of both risk factor and outcome, when the causal effect is constant (Scenario B2) and it when depends on U (Scenario B3)

		Scenario B2	Positive confounding (α_2 and β_2= 0.8), where β_1=0.5								
		Stratifying on collider, C					Stratifying on residual collider, C_0				
μ_1	μ_3	Proportion homogeneity rejected (%)	Median estimates Q1	Median estimates Q2	Median estimates Q3	Median estimates Q4	Proportion homogeneity rejected (%)	Median estimates Q1	Median estimates Q2	Median estimates Q3	Median estimates Q4
-0.5	-0.5	8%	0.07	-0.05	-0.03	0.06	6%	0.49	0.49	0.50	0.51
	0	5%	0.39	0.37	0.37	0.37	5%	0.51	0.49	0.52	0.51
	0.5	5%	0.60	0.66	0.65	0.61	5%	0.47	0.52	0.49	0.49
0	-0.5	6%	0.36	0.32	0.35	0.34	6%	0.50	0.51	0.51	0.49
	0	7%	0.48	0.50	0.52	0.48	6%	0.49	0.52	0.51	0.48
	0.5	5%	0.35	0.31	0.32	0.36	5%	0.50	0.50	0.51	0.50
0.5	-0.5	5%	0.54	0.54	0.58	0.54	5%	0.49	0.49	0.53	0.50
	0	6%	0.24	0.21	0.20	0.28	5%	0.48	0.52	0.49	0.52
	0.5	6%	0.10	0.01	0.02	0.10	5%	0.48	0.51	0.51	0.50
		Scenario B3	Positive confounding (α_2 and β_2= 0.8), where β_1=0.5+0.2U								
μ_1	μ_3	Proportion homogeneity rejected (%)	Median estimates Q1	Median estimates Q2	Median estimates Q3	Median estimates Q4	Proportion homogeneity rejected (%)	Median estimates Q1	Median estimates Q2	Median estimates Q3	Median estimates Q4
-0.5	-0.5	6%	0.09	-0.07	-0.07	0.00	9%	0.65	0.52	0.45	0.35
	0	7%	0.40	0.35	0.38	0.38	7%	0.52	0.50	0.51	0.48
	0.5	12%	0.45	0.59	0.69	0.85	11%	0.32	0.45	0.52	0.71
0	-0.5	7%	0.50	0.32	0.27	0.24	8%	0.69	0.50	0.43	0.36
	0	5%	0.43	0.46	0.51	0.57	4%	0.42	0.48	0.50	0.56
	0.5	9%	0.20	0.28	0.32	0.51	13%	0.34	0.47	0.53	0.70
0.5	-0.5	7%	0.66	0.58	0.58	0.41	7%	0.63	0.52	0.48	0.36
	0	5%	0.17	0.19	0.21	0.34	6%	0.36	0.49	0.55	0.64
	0.5	8%	-0.01	-0.03	-0.02	0.16	13%	0.31	0.46	0.53	0.69

Proportion homogeneity rejected represents the proportion of simulated datasets where the null hypothesis of homogeneity is rejected

Table F12. Scenario C1, results for positive confounding, adjustment approach, α1=0.1, considering a binary outcome Y

		Positive confounding (α_2 and β_2= 0.8)			
		Median estimate	**Type I error rate (%)**	**Median estimate**	**Type I error rate (%)**
μ_1	**μ_2**	*No adjust for collider*		*Adjust Y/G for collider*	
-0.5	-0.5	0.02	6%	-0.14	13%
	0	-0.01	5%	-0.11	8%
	0.5	-0.01	4%	0.02	5%
0	-0.5	-0.01	4%	-0.01	4%
	0	-0.01	4%	-0.01	4%
	0.5	0.00	5%	0.00	5%
0.5	-0.5	0.00	4%	0.02	6%
	0	0.02	5%	-0.09	7%
	0.5	-0.01	6%	-0.18	14%

Empirical Type I error rate represents the proportion of simulated datasets where the null hypothesis is not rejected

Table F13. Scenario C2 and C3, results for positive confounding, stratification approach, α1=0.1, considering a binary outcome Y, when the causal effect is constant (Scenario C2) and it when depends on C (Scenario C3)

		Scenario C2 **Positive confounding (α_2 and β_2= 0.8), where β_1=0.5**									
		Stratifying on collider, C					**Stratifying on residual collider, C_0**				
μ_1	**μ_2**	**Proportion homogeneity rejected (%)**	**Median estimates Q1**	**Median estimates Q2**	**Median estimates Q3**	**Median estimates Q4**	**Proportion homogeneity rejected (%)**	**Median estimates Q1**	**Median estimates Q2**	**Median estimates Q3**	**Median estimates Q4**
-0.5	-0.5	8%	0.17	0.16	0.15	0.22	6%	0.43	0.42	0.39	0.41
	0	6%	0.21	0.19	0.18	0.22	7%	0.37	0.38	0.40	0.36
	0.5	5%	0.39	0.37	0.38	0.36	4%	0.42	0.40	0.39	0.38
0	-0.5	4%	0.44	0.41	0.37	0.37	4%	0.43	0.39	0.37	0.37
	0	3%	0.41	0.42	0.41	0.38	4%	0.41	0.39	0.39	0.36
	0.5	5%	0.38	0.41	0.39	0.40	4%	0.40	0.41	0.39	0.40
0.5	-0.5	8%	0.28	0.29	0.36	0.38	7%	0.32	0.35	0.37	0.41
	0	6%	0.23	0.20	0.19	0.23	5%	0.38	0.38	0.42	0.39
	0.5	5%	0.27	0.16	0.13	0.23	5%	0.50	0.44	0.41	0.46
		Scenario C3 **Positive confounding (α_2 and β_2= 0.8), where β_1=0.5+0.2C**									
μ_1	**μ_2**	**Proportion homogeneity rejected (%)**	**Median estimates Q1**	**Median estimates Q2**	**Median estimates Q3**	**Median estimates Q4**	**Proportion homogeneity rejected (%)**	**Median estimates Q1**	**Median estimates Q2**	**Median estimates Q3**	**Median estimates Q4**
-0.5	-0.5	7%	-0.07	0.14	0.30	0.44	13%	0.01	0.37	0.57	0.70
	0	7%	0.04	0.18	0.26	0.44	10%	0.09	0.34	0.49	0.61
	0.5	6%	0.24	0.39	0.41	0.49	8%	0.19	0.37	0.47	0.54
0	-0.5	5%	0.27	0.38	0.49	0.62	5%	0.26	0.39	0.49	0.60
	0	6%	0.28	0.40	0.48	0.58	8%	0.27	0.40	0.49	0.58
	0.5	6%	0.31	0.39	0.52	0.61	6%	0.29	0.40	0.53	0.62
0.5	-0.5	7%	0.22	0.40	0.50	0.60	10%	0.23	0.43	0.58	0.66
	0	5%	0.14	0.23	0.33	0.50	7%	0.26	0.41	0.59	0.73
	0.5	6%	0.06	0.21	0.27	0.42	11%	0.22	0.50	0.62	0.78

Proportion homogeneity rejected represents the proportion of simulated datasets where the null hypothesis of homogeneity is rejected