Biomedicine & Pharmacotherapy 150 (2022) 112952

Contents lists available at ScienceDirect

Biomedicine & Pharmacotherapy

journal homepage: www.elsevier.com/locate/biopha

Review

Mendelian randomization in pharmacogenomics: The unforeseen potentials

Lubna Q. Khasawneh [a,1], Zeina N. Al-Mahayri [a,1], Bassam R. Ali [a,b,*]

[a] Department of Genetics and Genomics, College of Medicine and Health Sciences, United Arab Emirates University, United Arab Emirates
[b] Zayed Centre for Health Sciences, United Arab Emirates University, Al-Ain, United Arab Emirates

ARTICLE INFO

Keywords:
Mendelian randomization
Pharmacogenomics
Electronic health records
Instrumental variables
Interdisciplinary research

ABSTRACT

Mendelian randomization (MR) is an epidemiological method that uses genetic variants to proxy an exposure predicting its causal association with an outcome. It occupies a valuable niche between observational studies and randomized trials. MR applications expanded lately, facilitated by the availability of big data, to include disease risk causation prediction, supporting evidence of prior observational data, identifying new drug targets, and drug repurposing.

Concurrently, the last decade witnessed the growth of pharmacogenomics (PGx) research as a cornerstone in precision medicine. PGx research, conducted at discovery and implementation levels, resulted in validated PGx biomarkers and tests. Despite many clinically relevant PGx associations that could be translated into clinical applications, worldwide implementation is lagging far behind.

The current review examines the intersection zones between MR and PGx research. MR can provide supporting evidence that allows generalizing PGx findings supporting its implementation. Interchangeability, PGx research can fuel MR studies with libraries of genetic variants of validated biological relevance. Furthermore, PGx and MR exhibit a synergistic relationship in drug discovery that can accelerate identifying new targets and repurposing old drugs. Interdisciplinary research applied by PGx researchers, epidemiologists with MR experience, and data scientists' collaborations can unlock unforeseen opportunities in accelerating precision medicine acquisition.

1. Introduction

Since the beginning of the twenty-first century, Mendelian Randomization (MR) has gained mounting interest as a unique epidemiological approach to estimate causation in observational data [1]. MR uses the naturally occurring genetic variation as an instrument to measure associations [2]. For a long time, causal effects were inferred solely from extensive observational studies and randomized trials. Researchers are well customized to understand the observational studies' limitations and outcomes. One main limitation of these studies is the insufficient verification of confounders, leading to biased estimates of causal relationships [2]. In comparison, MR utilizes genetic variation, a life-long constant (except for the somatic genetic changes in tumors). Accordingly, the used variants are independent of potential confounding effects of the environment [1].

Simultaneously, at the opening of the 21st century, the field of pharmacogenomics (PGx) has gained momentum as a cornerstone in the precision medicine evolution. The scope of pharmacogenomics (PGx) encompasses applying genomic and molecular technologies to drug design, evaluation, and assessment [3]. The PGx biomarkers are found in genes involved in drug pharmacokinetics and pharmacodynamics [4]. Eventually, PGx helps determine which drugs and dosages are best suited for specific individuals based on particular biomarkers. In this context, PGx testing enables identifying problems with current treatments that could lead to poorer outcomes for patients or increase adverse events [5].

Different study designs generate the PGx-guided therapy guidelines. Observational studies usually lead to recognizing PGx biomarkers, with the frequent employment of case-control design to examine a PG-x biomarker and a given drug outcome association. Confounders frequently hamper the cohort and case-control studies, and PGx studies are not an exception. While ideally, randomized controlled clinical trials (RCTs) are needed to evaluate the implementation of PGx testing in clinics [6]. However, RCTs are not always feasible in the real world. Ethical considerations and the high cost of these studies are among the reasons for their inapplicability [7].

* Correspondence to: Department of Genetics and Genomics, College of Medicine and Health Sciences, United Arab Emirates University, P.O. Box: 15551, Al-Ain, United Arab Emirates.
E-mail address: bassam.ali@uaeu.ac.ae (B.R. Ali).
[1] Equally contributed

https://doi.org/10.1016/j.biopha.2022.112952
Received 7 March 2022; Received in revised form 7 April 2022; Accepted 8 April 2022
Available online 13 April 2022

Accordingly, the extent to which MR methodology might intersect with PGx research is worth investigating. Throughout the current review, we will evaluate the intersection areas between the unique MR approach and PGx, and the potential opportunities. Herein, we summarize the emergence of MR and its different designs and applications. Then we provide a brief description of PGx research and the difficulties hampering its findings. Finally, we illustrate the possible contributions that MR can have to the PGx research and vice versa.

2. MR studies; definitions and different designs

MR can be considered a "natural" RCT, where the randomization has occurred randomly at conception according to Mendelian laws of segregation and independent assortment [1]. Besides ensuring unbiased randomization, genetic variation is fixed since conception and precedes the measured variable, enabling the reverse causation effect to be overcome. In other words, there are no chances of the effect causing biased randomization towards any of the study arms assigned according to constitutional alleles [2]. However, evidence derived from MR depends on the satisfaction of its assumptions and avoiding the biased selection of variables, making it at a lower strength than RCTs. Nevertheless, being less susceptible to confounding effect and reverse causality places it above observational studies, including case-control ones [2]. Table 1 compares MR studies with RCTs and case-control studies and summarizes their fundamental differences.

Conducting MR starts by identifying an intermediate variable or a biomarker associated with the disease risk. The second step will be selecting a genetic variation known to affect the biomarker level with strong replicated evidence. Finally, the genetic variant, which is called then the instrumental variable (IV), is tested for association with disease risk. If the genetic variant is associated with the disease, it is inferred that the biomarker levels cause the disease [8].

As with other statistical methods, MR relies on several notions. The selected IVs should fulfill three assumptions; they should be strongly associated with the studied risk factor or biomarker (the relevance assumption). Secondly, there should be no association with confounders (the independence assumption). Finally, the relation between these variables and the outcome of interest should be solely through the risk factor and not any other pathways (the exclusion restriction assumption) [2]. MR assumptions are illustrated in Fig. 1.

The first published MR studies were performed using a one-sample design. One-sample MR, also called single-sample MR, measure the instrument, the exposure, and the outcome of interest in the same group. In later studies, two-sample MR proved feasible. In the latter design, the exposure and outcome are measured in different groups or partially overlapping ones. The two-sample MR surpasses the one-sample MR by using the publicly available GWAS data and avoiding the high cost and difficulty of measuring the exposure and the outcome in the same group [9]. With the increasing accessibility to GWAS-derived data, robust associations between genetic variants and exposures of interest are available now for researchers. Accordingly, the potential of the two-sample MR approach is rapidly evolving, with costs much lower than RCTs [10].

Another design of MR is known as "bidirectional MR." In this approach, the analysis involves examining whether the exposure causes the outcome, and vice versa, whether the outcome causes the exposure. Moreover, when an intermediate trait is suspected to be a causal mediator between the exposure and the outcome, a two-step MR is used. The two steps include examining the exposure-potential mediator association using first the exposure's IVs, then the mediator's ones [9].

Frequently, genetic variants exhibit horizontal pleiotropy through correlation with multiple phenotypes. In this case, using a single variable MR can result in a false causality conclusion. To overcome this limitation, multivariable MR can be conducted using instruments associated with multiple exposures to collectively estimate the causal effect of independent risk factors on the outcome. On the other hand, factorial MR can assess the impact of the combined occurrence of two or more risk factors as disease causals [9].

Table 1
Comparison of the Mendelian randomization, case-control studies, and randomized controlled trials.

	Mendelian randomization	Case-control studies	Randomized Controlled Trials
Type of study	Analytical studies to estimate the causal effect from observational data	Observational studies (subjects are observed to determine the outcome)	Interventional studies to assess the impact of an intervention
The study participants	The participants' data is usually derived from large databases or previous studies and allocated into two arms according to the carrier status of the selected genotypes	The participants are selected from the population (or from available databases) and allocated into two arms: the disease group (case) and a chosen matching group of disease-free participants (controls)	Participants are recruited to the study according to pre-defined inclusion criteria and randomly allocated into two arms (e.g., treatment versus placebo). Specific populations are frequently excluded (e.g., pregnant women, the elderly)
Randomization method	Subjects are assigned randomly on their genetic variants (e.g., carriers of risk allele versus carriers of protective allele)	Subjects are assigned in the control arm to match the cases in factors other than the exposure under question (no randomization)	Subjects are assigned randomly on pre-defined arms.
Confounding effect	Confounders are equal in both arms and partially controlled	Confounding effects cannot be controlled and can lead to bias	Confounders are similar in both arms and strictly controlled
Reverse causation	Low susceptibility to reverse causation	The disease state can affect the biomarker's level causing reverse causation	Not applicable
Life-long exposure	Provide estimation of the life-long exposure as the variants are inherited since birth	Estimation of life-long exposure is not possible	An estimate of life-long exposure is not possible
Causality	Can give evidence of causality (given that all assumptions are fulfilled). Results should be interpreted in the context of existing evidence from other sources.	It cannot establish causality but results in associations	Can provide the most substantial evidence of causality.

3. Common applications of MR

Initial applications of MR focused on estimating the causal relationships between a specific environmental exposure and a disease. Later studies employed MR in other applications, resulting in significant outcomes [1]. Table 2 illustrates recent examples of the different MR applications, while the following paragraphs elaborate on some important studies under each objective.

3.1. MR identifies disease risk factors

Identifying disease risk factors was the earliest application of MR.

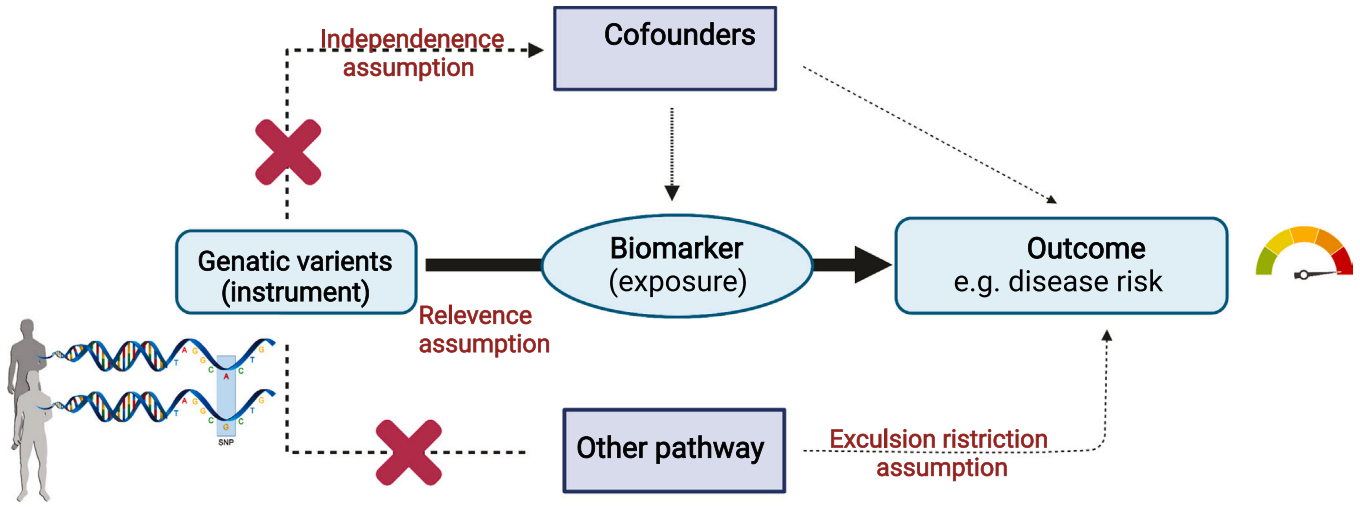

Fig. 1. : Mendelian randomization concept and assumptions.
(Created with BioRender.com).

Table 2
Recent examples of MR applications.

MR Application	Genetic variant (instrument)	Exposure	Outcome	Main finding	References
Identifying disease risk factors	*ADH1B* variant	Alcohol consumption	stroke risk	A positive causal association between alcohol consumption and stroke development was confirmed	[11]
Identifying drug target	Loss of function variants in *PSK9*	LDL-C	Evolocumab (PSK9 antibody) as a drug target for familial hypercholesteremia	Evolocumab reduced LDL-C levels by 28% and declined CHD risk to 88%.	[12]
Giving early predictions of ongoing RCTs	Genetic variants of ACLY inhibitor	LDL-C	Bempedoic acid, an oral ACLY inhibitor, in lowering cholesterol levels	Bemedoic acid reduces LDL-C levels.	[13].
Supporting earlier studies evidence	16 genetic variants that are found from GWAS to be associated with AF	AF	Effect of AF on brain volume	AF affects low grey matter brain volume but not the white matter	[14]
Drug repurposing	genetic instruments that proxy 1263 actionable proteins targeted by approved drugs	Actionable proteins	reduce the risk of COVID-19 hospitalization	type I interferons (IFNAR2) and ACE2 modulators (ACE2) suggested being part of the early management of COVID-19	[15].
Predicting adverse events	*PSK9* variants	LDL-C	PSK9 inhibitors and incidence of type-2 Diabetes	PSK9 inhibitions, which proxy reduction in LDL-C, modestly increase the risk of type-2 diabetes.	[16].
Providing evidence through the development of "informatics consult."	Genetic variants that affect circulating vitamin K quantities	AF in patients with liver cirrhosis	Efficacy of warfarin for stroke prevention	MR and two other recourses provided evidence that was absent in the RCT. Warfarin is efficient in reducing stroke risk in patients with AF who are comorbid with liver cirrhosis	[17]

ACE2: Angiotensin-Converting Enzyme 2, ACLY: ATP citrate lyase, ADH1B: alcohol dehydrogenase B, AF: atrial fibrillation, CVD: Cardiovascular Disease, CHD: Coronary Heart disease, COVID-19: Coronavirus Disease-19, IFNAR2: Interferons, GWAS: genome-wide association studies, LDL-C: Low-Density Lipoprotein Cholesterol, MR: Mendelian randomization, RCT: Randomized Controlled Trials, PSK9: proprotein convertase subtilisin-Kexin type 9.

Multiple examples can be found in the literature. A classic example includes MR studies applied to examine the association between lipoprotein levels and cancer development. Numerous observational studies reported an association between low cholesterol levels and cancer development. However, there were doubts that the reported association originated from reverse causality (i.e., cancer patients have abnormally lower cholesterol levels, so cancer caused low cholesterol levels and not the opposite) or confounding (other factors like smoking or age were not considered). Accordingly, MR was suggested to resolve this confusion. Variants in the gene encoding for apolipoprotein E (*APOE*), which is involved in the clearance of lipoproteins from plasma, were used as IV. This investigation included 2913 elderly participants and found no causalty between life-long lipid levels, proxied by *APOE* variants, and increased cancer risk [18].

Similarly, the causation between C-reactive protein (CRP) levels and ischemic heart disease (IHD) development was evaluated. The life-long increase in CRP levels was proxied by four combined CRP polymorphisms. The results show that genetically inherited elevation in this inflammatory mediator is strongly associated with increasing the levels of CRP but not with increased risk of IHD. Depending on MR analysis of data produced from four large studies that included more than 38,000 participants, the authors concluded that the increased IHD rates associated with CRP high levels observed in epidemiological studies might not represent a causal relationship [19].

More recently, MR was used to examine the controversial association between alcohol use and cardiovascular diseases (CVDs). A positive association between alcohol consumption and stroke incidence was reported in a recent large Chinese cohort. The IVs used in this study included a variant in the alcohol dehydrogenase (ADH) gene, namely rs1229984 in *ADH1B*, which is the primary enzyme in alcohol degradation. The other variant was an aldehyde dehydrogenase variant (rs671 in *ALDH*-2), a common variant in Asians that leads to slow aldehyde

breakdown and accumulation, resulting in severe discomfort and lower alcohol consumption. The outcomes of this study, derived from the genetic epidemiology of more than 500,000 individuals, showed that alcohol intake uniformly increases blood pressure and stroke risk [11].

3.2. MR identifies drug targets

As described earlier, MR was introduced as a strategy to provide evidence for causal relations between a biomarker and a disease. When a biomarker is a causal factor in disease pathogenesis, then the genetic factors that affect this biomarker will affect the disease progression. Hence, when MR proves the causation of biomarkers, these become plausible targets for drug design. Nevertheless, few examples can be found in the literature about prospective studies that used MR for drug development [8].

This application is demonstrated by identifying proprotein convertase subtilisin/Kexin type 9 serine protease (PCSK9) as a plausible drug target. PCSK9 is a glycoprotein that degrades low-density lipoprotein cholesterol (LDL-C) receptors (LDL-R), responsible for removing LDL-C from circulation. Following this degradation, the cellular uptake of LDL-C decreases, and higher levels will be circulating in the blood. Accordingly, people with constitutional loss-of-function variants at *PCSK9* will have a life-long reduction in serum LDL-C levels. Interestingly, the genetic-based analysis revealed that the 15–28% decrease of mean LDL-C in subjects with *PCSK9* loss of function variants conferred coronary heart disease risk decline by 50–88%, respectively [20]. This notion, derived from MR studies [21] and supported by RCTs, nominated PCSK9 antibodies as a plausible drug target and led to the development of evolocumab as the first monoclonal PCSK9 antibody [8]. Evolocumab was approved in 2015 to reduce LDL-C levels in hyperlipidemic adults, then in 2017 to reduce cardiac events in patients with atherosclerosis. More recently, this drug proved its efficacy and safety in pediatrics with familial hypercholesteremia [12].

3.3. MR giving early predictions of ongoing RCTs

By utilizing the existing genomic databases, MR can predict the outcomes of clinical trials in advance. The work of Holmes and colleagues demonstrates a typical example. They examined the probable effect of lowering the secretory phospholipase A2 (sPLA2-IIA) on cardiovascular events by utilizing a variant in *PLA2G2A*, namely rs11573156, as an IV. The selected IV affects the mass of circulating PLA2-IIA. However, in their MR-based meta-analysis of 19 populations, Holmes and coworkers found no association between the chosen IV and major vascular events. Accordingly, they concluded that therapeutic approaches to reduce the sPLA2-IIA mass are unlikely to show efficacy in preventing CV events [22]. This prediction was reported before announcing the adverse outcomes of the sPLA2 inhibitor, varespladib, on cardiovascular outcomes from the large RCT that was still ongoing then, before early termination due to futility and possible harm [23].

In contrast, MR gave supporting evidence for using bempedoic acid, an oral ATP citrate lyase (ACLY) inhibitor, in lowering LDL-C levels before the conclusion of the largest RCT evaluating the same agent. In their work, Ference and colleagues validated ACLY inhibition as a genetic target with a mechanism similar to statins. The authors found a decrease in cardiovascular events with ACLY inhibition, making it a suitable therapeutic target [13].

These examples illustrate how MR can provide early prediction of RCTs outcomes. The feasibility of using MR for this purpose is currently more than ever, with the unprecedented increase in the genomic data derived from large GWASs.

3.4. MR supporting earlier studies evidence

A classic example of this application is demonstrated by studies supporting the association between high levels of LDL-C and CVDs risk. Multiple genetic variants in the LDL-receptor gene or other genes identified are associated with LDL-C levels. MR studies that utilized these variants as IVs confirmed the effectiveness of lowering the LDL-C levels in reducing the CVDs risk [24,25]. Nevertheless, these studies were conducted several years following the introduction of the cholesterol-lowering drugs, statins, into the clinics and following multiple RCTs that established statins' efficiency in protecting from CVDs. In this case, the latter MR results were confirmatory to older RCTs [8].

Another example that showed MR as a robust tool for validating previous research findings is the multivariable MR analysis performed to estimate the causal effect of atrial fibrillation (AF) on brain volume. Park and colleagues used 16 genetic polymorphisms derived from a metanalysis of previous GWASs as strongly associated with AF, with independent brain volume measures from the UK Biobank, in a two-sample MR design. The authors concluded that AF disease affected the lower grey matter volume of the brain but not the white matter volume. The detected causation, which supported previous observational evidence, was found to be possibly mediated by the effect of an ischemic stroke [14].

3.5. MR in drug repurposing

MR strategies can suggest new uses for old drugs. For example, multiple prospective observational studies concluded that high circulating concentrations of interleukin 6 (IL-6) are associated with an elevated risk of coronary heart disease. This effect is caused by an inflammatory process mediated by the IL6 receptor (IL-6R). MR-based meta-analysis of 40 studies that used an IL-6R variant as an IV reported a causal role of IL-6 in coronary heart disease development. Given this association, repurposing of Tocilizumab, an IL6R monoclonal antibody licensed for rheumatoid arthritis, was suggested [26].

Moreover, some genes are recently identified in the literature as actionable druggable genes. This notion is used for genes encoding the protein targets of licensed drugs or drugs in the clinical phases of development. MR studies carried out using variants in these genes can identify repurposing opportunities for the approved/clinical phase drugs [15].

A significant application of the latter approach is illustrated by the recent work of Gaziano and colleagues within the COVID-19 body of research. The authors used genetic instruments that proxy 1263 actionable proteins targeted by approved drugs. The association between these IVs and COVID-19 hospitalizations (7554 hospitalized patients with COVID-19 and more than one million controls) pointed out three proteins to be co-localized with hospitalization. Further assessments, by phenome-wide analysis and pathway enrichment, prioritized drugs targeting type I interferons (IFNAR2) and ACE2 modulators (ACE2) for the early management of COVID-19 [15].

3.6. MR predicting adverse events

Adverse drug events are a significant health concern ranked among the top ten leading causes of death [27]. In the post-marketing phase, adverse events are identified through reporting systems that rely on health care professionals reporting. Strong signals from these systems are later investigated through meta-analysis of RCTs and observational studies. However, the reporting system suffers from several biases, and proving causality from observational studies has its limitations, as discussed earlier [28].

Representative studies of this MR application can be demonstrated by studies that explored statins and their induced increased risk of type-2 diabetes. Multiple MR studies were conducted to examine the causality of statins on hyperglycemia. This adverse event was primarily recognized by a meta-analysis of the RCTs [29]. Swerdlow and coworkers hypothesized that statin's mechanism of action might be behind this effect. To investigate this hypothesis, they applied MR analysis utilizing an IV from the gene encoding for the enzyme

3-hydroxy-3-methylglutaryl-CoA reductase (HMGCR). The latter is the target for statins, and a specific variant in it (rs17238484) was found, through a large GWAS, to be associated with lower LDL-C concentration. The results demonstrated increased plasma insulin and glucose concentrations for each variant allele. The authors concluded that inhibiting HGMCR by statins can partially explain the increased risk of diabetes reported with statins [30].

Using a similar approach, Schmidt and coworkers used MR to examine the effect of lowering LDL-C on type-2 diabetes development. They found an association between the *PCSK9* variants, which proxied pharmacological LDL-C reduction by *PCSK9* inhibition, and a modestly higher risk of type-2 diabetes. However, they used PCSK9 variants as IVs; accordingly, the outcomes reflect the adverse effects of PCSK9 inhibitors rather than statins in this case [16].

Collectively, both studies pointed out that lowering LDL-C increases the risk of type-2 diabetes. This conclusion was later contradicted by the work of Smit and colleagues [31], which will be discussed in a succeeding paragraph. Worth mentioning MR is limited to predicting the adverse events related to the drug's mechanism. In contrast, the Off-target effects, which are not associated with the drug's mechanism of action, cannot be predicted by MR [13]. Nevertheless, the previous studies exemplified how MR can support evidence of drug-induced adverse events.

3.7. MR provides evidence through the development of "informatics consult"

"Informatics consult" is a recently proposed evidence-based clinical decision support tool. This approach aims to provide the clinician with on-demand evidence guidelines based on the analysis of aggregated electronic health records (EHR) data. Traditionally, RCTs are the gold standard to support clinical treatment guidelines. However, generating evidence from trials when patients complain of comorbidities is not possible, as most RCTs exclude patients with comorbidities. Accordingly, the "informatics consult" aims to cover this gap by providing evidence-based guidelines for such cases with incomplete information to support the decision-making [17].

Recently, Lai and colleagues evaluated the "informatics consult" approach using a common controversial clinical decision as an example. The utilized model is the decision to prescribe warfarin for stroke prevention following AF diagnosis, one common warfarin indication, in patients with liver cirrhosis, a common warfarin contraindication. To evaluate the efficacy of "informatics consult" in this scenario, the authors analyzed four sources of evidence: RCTs, a meta-analysis of observational data, trial emulation (i.e., real-life data from EHR is used to mimic an RCT), and MR approach. For the MR approach, the selected IVs were those variants affecting the circulating amount of vitamin K. Given that warfarin is a vitamin K antagonist, the selected IVs will proxy the warfarin use, so the causal relationship between warfarin use and stroke risk is evaluated. Interestingly, the authors found a complete absence of RCTs or trials reporting results of patients with AF and cirrhosis, as cirrhosis is a common exclusion criterion from RCTs. In comparison, MR and the other two sources showed, with a broad concordance, the efficacy of warfarin in reducing stroke risk among patients with both cirrhosis and AF [17].

Herein, MR compensated for the infeasibility of conducting RCTs to generate guidelines in cases of comorbidities. Eventually, it served as a valuable method to validate the data-informed clinical decision-making known as informatics consult. Such tools are envisioned to bridge the gap between evidence and practice. Despite that changing clinical practice guidelines depending on MR studies is still far from being achieved, MR analysis strengthens the confidence in artificial intelligence-depended clinical decision support tools. This confidence may eventually lead to a better acceptance and adoption of guidelines generated through the novel and unconventional pathways.

4Difficulties Hindering MR

Despite the successful findings from applied MR studies illustrated in the previous examples, inherent difficulties hinder the utilization of MR in many cases. In the following paragraphs, we highlight some of the complications faced during the conduct of MR studies.

4.1. Pleiotropy

As described earlier, MR gives positive results when the genetic variant known to influence the exposure also affects the outcome. Some researchers have termed this vertical pleiotropy which is considered an essential component of MR. In contrast, horizontal pleiotropy, also known as biological pleiotropy, is a major limitation that should be addressed during the MR study design. The latter describes the cases where the genetic variant affects the outcome in a pathway other than the exposure. It can also occur when the selected IV is in linkage disequilibrium with a causal variant for another trait [32].

As an example, horizontal pleiotropy hindered the success of an MR study using *CETP* variants. Cholesteryl-ester transfer protein (CETP) facilitates transferring of cholesteryl-ester and triglycerides between LDL and HDL. Accordingly, inhibition of CETP, which maintains the cholesterol in HDL particles, formed a probable drug design target. Nevertheless, examining the SNPs occurring in *CETP* and following MR approaches to evaluate the effect of these naturally occurring variants on the risk of CVDs failed to prove efficacy. This failure was suggested to be a consequence of the pleiotropic nature of CETP. In other words, the weak association between *CETP* variants and coronary heart disease risk is thought to be due to the effects of CETP on pathways other than increasing HDL-C levels [8].

Several tools are developed to recognize the horizontal pleiotropy effects in order to improve the causal effect inference reliability. Such methods, reviewed elsewhere [32], can model these effects and, with the help of machine learning and the recent massive data available, can reduce them and increase the consistency of MR findings.

4.2. Canalization

The term refers to the process in which the genetic polymorphism effect is buffered through the expression of another polymorphic genotype during fetal development as a developmental compensation. In other words, either another gene with an identical or similar function (i.e., gene redundancy) or another metabolic pathway buffer against the effect of a genetic polymorphism. Accordingly, a significant caveat will arise when attempting to use these polymorphisms in MR as IVs to proxy an outcome, while indeed, the outcome has been compensated by another biological process [33].

However, it is not clear to what extent MR results are affected by canalization. It was suggested that exposures that occur in adulthood, like alcohol consumption, are unlikely to have been compensated for during development. In this context, *ALDH-2* variants can be considered reasonable proxies for alcohol consumption that are unlikely to be affected by canalization [34].

Nevertheless, canalization is still considered as an unavoidable limitation in MR studies that can explain the lack of genetic evidence on expected associations. For instance, a recent study evaluated the effect of circulating levels of interleukins-1 family members and lung cancer in a large two-sample MR. The findings demonstrated low genetic evidence of an association between interleukin-1Beta (IL-1β) and lung cancer, in contrast to previous RCTs findings. The authors suggested canalization as one unavoidable limitation that might partially explain the unexpected finding of their MR analysis [35].

4.3. Interpretation of the results

What is beyond MR predictions, and how will MR results be

translated into clinical practice? In which disciplines are MR investigations required? Can we rely on MR results to confirm causality relationships in diseases and drug targets? These questions and many others are still vague since there are no standard guidelines or recommendations to interpret the causal effect estimated from MR associations which hinders their full utilization.

Burgess and coworkers suggested applying "Bradford Hill criteria" to judge the plausibility of IV assumptions before considering the causality resulting from MR. Hill's criteria for causation are a widely accepted checklist of nine aspects that validate the instrumental variable assumptions and support the causal-effect evidence. Hill's criteria suggest that genetic variants should have a well-established biological role rather than only a statistical relevance for a well-justified MR analysis assessment. Accordingly, applying these criteria will favor IV derived from candidate gene studies over those from GWASs. The authors clarified that causation resulting from MR studies that used IVs with statistical significance only, like those mainly derived from GWAS results, will provide less reliable causations [36].

Interestingly, Bradford Hill criteria were used in the MR analysis applied by the "COVID-19 host genetics initiatives". The MR, in this case, was used to support the evidence of causation/no-causation of factors suggested increasing the susceptibility for severe acute respiratory syndrome coronavirus 2 (SARS-COV2) infection and its associated complications. The results revealed a causal role of smoking and body mass index in severe COVID-19. In contrast, MR outcomes could not support the causation of type- 2 diabetes in COVID-19 severity [37].

Hill's checklist can be used as an aid to infer causation. Undoubtedly, other approaches still need to be developed to increase the confidence in MR predictions and address their translation questions.

5A brief description of PGx Research and its challenges

PGx research is conducted to answer the following questions: Which genes are involved in drugs' mechanisms? Which pathways are involved in propagating the drug's effect? How to explain and characterize "off-target" adverse drug effects through genetic information? And how the gene-drug interaction information can be utilized in clinical decisions? [38]. PGx research is applied under two broad classes: PGx discovery and PGx translational research to answer any of the previous questions. The discovery research aims to find PGx associations either within genes with prior knowledge of their involvement in pharmacology, like genes involved in drug metabolism or genes encoding drug receptors, or by exploring the genome. The latter investigations employ agonistic approaches, like GWAS or whole exome or genome data, to assess the genetic variants behind various drug response parameters. Advances in research techniques fueled the PGx discovery efforts with an exponentially growing list of rare variants with reliable estimates of their impact on the drug response [39].

A significant challenge in applying GWAS for PGx associations discovery is the large sample size inherently needed in GWAS. In comparison, most drug adverse events are rare and unpredictable. This implies the requirement of longitudinal studies including large populations. Prospective studies are usually not large enough to illustrate rare events, while retrospective studies data collection can be brutal due to high morbidities and mortalities among the study cohort [40].

On the other hand, translational PGx research utilizes the findings from the discovery phases to examine their implementation in the clinic. The implementation usually passes through different stages; following the discovery, a genetic test and intervention should be developed. Such implementation will lead to evidence-based guidelines, which eventually require a real-world evaluation of their impact on the healthcare system [39].

Studies applied across these stages are crucial to moving the PGx testing into standard care. One commonly used design in the earlier stages of evaluating pharmacogenetic testing is observational PGx studies. These studies frequently assess gene-drug interaction using a population-based case-control design. The study subjects are divided into subjects with positive treatment outcomes (i.e., good response or no toxicity) and those with a negative outcome arm. The two arms are then genotyped to create gene-drug case and control groups. Despite the ease in performing such design and its ability to provide direct associations, they suffer from confounding effects, the need for large sample sizes, difficulties in defining the cases and controls, and other caveats that may lead to bias [41].

Despite the growing list of clinically relevant PGx biomarkers recognized to improve patient care, implementation in the clinic is lagging. There are multiple reasons behind this lag; difficulties in interpretation, scarce education about PGx among clinicians, and limited available evidence to support the PGx-testing clinical utility, among others [42]. Accordingly, approaches that can circumvent any of these hurdles continuously need to be developed and evaluated.

6. Intersection points between PGx and MR

As described earlier, genetic variants that explain a modifiable risk factor variability are used in MR to explain the causal effect of this risk factor on an outcome. In comparison, can the "exposure" to a specific medication be considered an "environmental exposure/risk factor," and can an approach similar to MR be used to measure the causality relationship between this exposure and a defined outcome?

To answer this question, the MR assumptions should be validated. The PGx variant should be strongly associated with the risk factor, i.e., being exposed to the medication to fulfill the relevance assumption. The same variants should not be associated with a confounder to satisfy the independence assumption. Finally, the exclusion restriction implies that the relationship between genetic variants and the outcome should be through this medication, defined as exposure only and not any other pathways.

For example, the antiplatelet clopidogrel is a prodrug primarily activated by CYP2C19. Accordingly, the *CYP2C19* variants can be used as IVs to measure the causality between clopidogrel exposure and an outcome like stroke incidence. Loss of function (LoF) variants in *CYP2C19* are strongly associated with exposure to clopidogrel (relevance assumption), without any substantial evidence supporting the effect of other enzymes on this activation process (the independence assumption). There is no evidence supporting that clopidogrel will cause strokes through mechanisms other than its inefficient activation by CYP2C19 due to gene's LoF variants (exclusion restriction assumption). The critical difference between IVs in the PGx context and conventional MR studies is that the gene variants affect the exposure (i.e., the drug) *once* taken [43].

In recent work, Pilling and colleagues analyzed stroke and myocardial infarction (MI) incidence in more than 7400 patients from the UK-Biobank who were prescribed clopidogrel for more than two months. The same *CYP2C19*-LoF variants' carriers were compared to non-carriers. The authors found that patients prescribed clopidogrel who carry *CYP2C19*-Lof variants had a significantly increased risk of stroke compared to non-carriers (1.9% and 1.6%, respectively), and similar figures were reported for MI incidence (26.4% and 24.1%, respectively). In the same study, the authors confirmed that the effect of the CYP2C19-Lof genotype on stroke incidence was specific to clopidogrel and not through any other biological pathways. This conclusion was based on the analysis of stroke events in 198,868 UK-Biobank participants who were not exposed to clopidogrel, in which the CYP2C19-Lof was not associated with stroke events [44].

The same research group aimed to find if the association between CYP2C19-Lof and stroke incidence stays still in the general population, i.e., if such findings can be generalized at the population level, including those never treated with clopidogrel. Bowden and colleagues reanalyzed the clopidogrel-*CYP2C19* dataset from the UK Biobank to answer this question. They examined a proposed framework that estimates the "Genetically Moderated Treatment Effect," or GMTE, in the general

population, i.e., treated and untreated with clopidogrel. The analysis illustrated that when PGx variants assumptions are partially violated, MR can yield a consistent estimate of GMTE and allows for calculating treatment effect reduction in patients with particular variants. The proposed framework, referred to as triangulation within a study (TWIST) framework, illustrated that the *CYP2C19*-Lof carriers experienced a 0.28% increased risk of stroke when treated with clopidogrel. This result was supported by a considerable sample size (the UK biobank), which is hard to replicate in an RCT. Accordingly, this study has shown how PGx and MR can work side by side to estimate GMTE and aid in decisions related to public health, which are challenging to make using the conventional RCT designs [43]. To the best of our knowledge, this research was the first attempt to embrace the PGx findings with the MR approach in a large dataset to answer a question related to public health. The following paragraphs illustrate more examples of the intersection points between PGx and MR approaches, also illustrated in Fig. 2.

Different applications of MR are depicted (dark blue circle), and the general research workflow of PGx from discovery to implementation is illustrated (green circle). GWAS (light blue circle) is heavily utilized, among other methods, in PGx discovery studies and is one key source of genetic variants used as instrumental variables in MR. The suggested interactions between MR and PGx are exemplified in three points (1) Data from EHR (grey circle) can be analyzed through MR to estimate the GMTE which, if proved significant, supports the implementation of PGx testing at the population level. (2) Genetic variants discovered in PGx (PGx variants) can be used in MR applications as IVs with biological relevance. (3) Both MR and PGx can support and accelerate drug discovery research. (Created with BioRender.com) *AE; adverse effects, EHR; electronic health records, GMTE; genetically moderated treatment effects, GWAS; genome-wide association studies, IVs; instrumental variables, MR; Mendelian randomization, PGx; Pharmacogenomics, RCT; randomized control trials.*

6.1. MR as a tool to support PGx implementation

Higher quality evidence is needed to speed up the PGx implementation. As illustrated earlier in Bowden's and colleagues' work, MR analysis can aid the PGx studies outcomes to provide complete answers to the effect of PGx testing on the health systems [43]. Leveraging the observational databases through MR analysis can address the demand for supporting evidence in PGx. Moreover, MR can provide more solid associations between PGx variants and adverse outcomes in the medical records of a drug-exposed group of patients. Such information can be used further to estimate the effect of PGx guided therapy [45].

Such an approach has been rarely utilized in the literature. Tornio and colleagues exemplified it in a small real-world study. The authors used electronic medical records data to extract the stroke re-occurrence rates. These rates were examined among clopidogrel users who were randomized, through MR, to either possess the *CYP2C19** 2 allele or not.

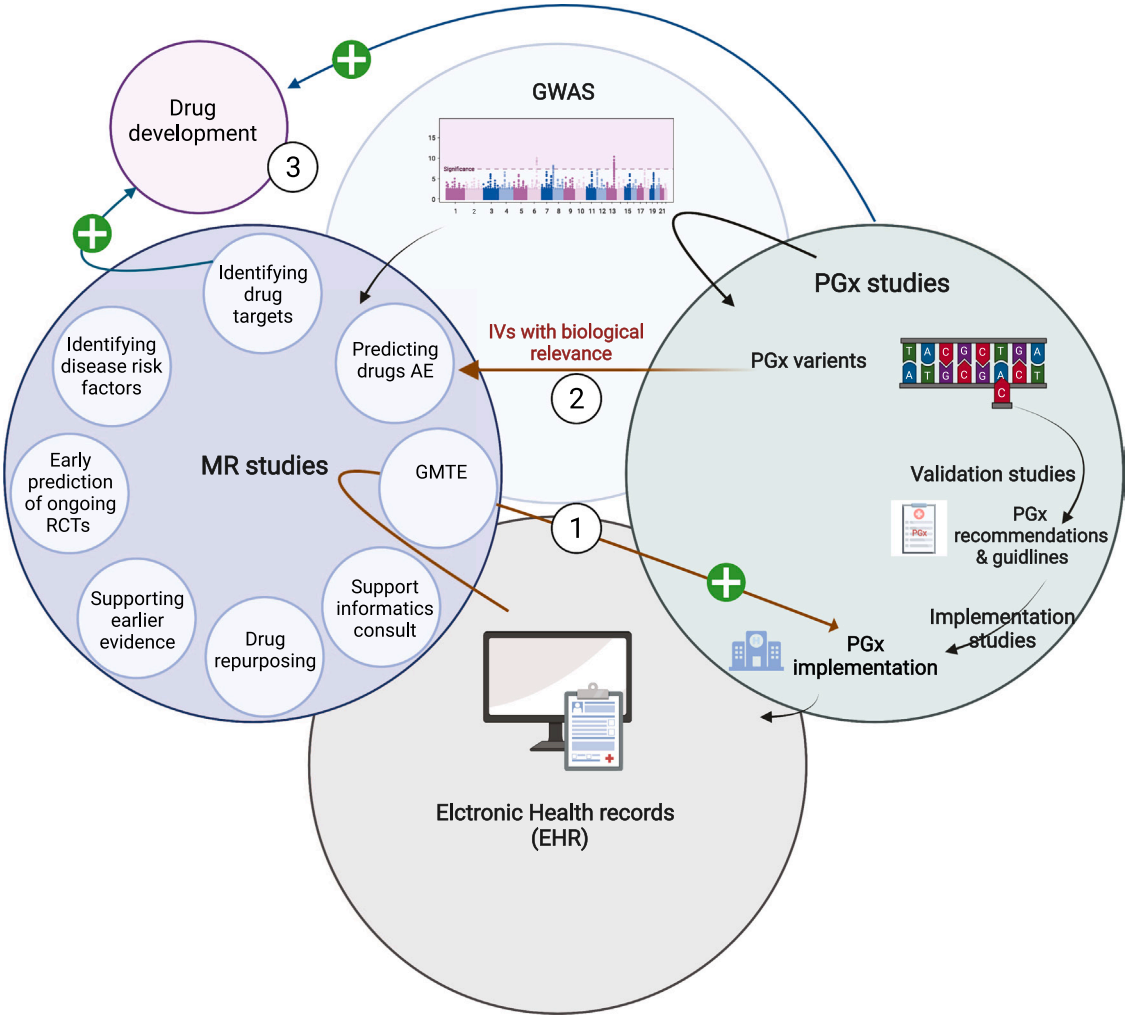

Fig. 2. : The suggested interaction paradigm between pharmacogenomic research and Mendelian randomization approaches. (Created with BioRender.com).

The latter allele is one common *CYP2C19*-Lof allele. It was found that almost 25% of stroke patients treated with clopidogrel remain significantly at higher risk of a recurrent stroke due to their *CYP2C19** 2 carrying status. However, this study had limitations, including the small sample size and analyzing a single genetic variant. Nevertheless, it is one of the few studies that utilized such an approach and illustrated the potential value of PGx testing in the real world rather than the research setting [46].

6.2. Confirmed PGx associations can fuel studies using MR with robust IVs

Indeed, the relation between PGx research and MR-based studies is bidirectional. In recent years, GWAS investigations have been increasingly used to answer questions of PGx research. Such studies provide strong associations between genetic variables and specific drug-related phenotypes. At the same time, they form rich IVs generators for many epidemiological MR-based studies.

For instance, the association between statins and type-2 diabetes was recently re-examined by Smit and coworkers. Previous MR studies, discussed earlier in this review, which examined the association between cholesterol-lowering drugs and the development of type-2 diabetes as an adverse event, used genetic variants proxying life-long LDL-C levels. Instead, Smit and colleagues employed the findings of the most extensive PGx study on statins to designate variants to proxy statin response. The selected IVs were used to examine the causal effect of statin response on type-2 diabetes in a two-sample, bidirectional MR approach. Besides demonstrating no evidence of the suggested association between statin and type-2 diabetes development, this study illustrated how PGx data could provide a robust library of IVs for MR studies [31].

Nevertheless, the authors commented that their reported lack of association between statins and type-2 diabetes might be partially related to including non-statin users in the dataset, which might have diluted the effect estimate. Accordingly, while they encourage using IVs derived from PGx studies, they recommend that the assessment of causal effects, in this case, should be examined in populations composed solely of drug users [31].

6.3. PGx and MR in drug discovery and development

Identifying drug targets is a common outcome of research utilizing MR and, at the same time, is a primary target of PGx research. PGx-based associations can identify druggable pathways and potential drug targets. PGx research can identify variants that can potentially affect dose, response, and/or adverse effects following initial trials and animal models. Moreover, utilizing PGx information at the early stages of clinical trials enables the exclusion of "non-respondents," ensuring the drug's passage through the clinical trial into the market [38]. Indeed, there has been a consensus on the utility and indispensability of PGx information in all stages of drug development. Stratification of participants according to PGx findings can inform the later-phase design and permit developmental decisions at shorter intervals [47].

Pharmaceutical companies are increasingly collecting DNA samples from clinical trials participants. Nevertheless, conducting PGx evaluation during the different phases of drug development is restricted by regulations and the complex set of requirements of regulatory bodies across the globe. Accordingly, impeding PGx assessments during drug development is still limited despite its well-recognized benefits. One suggested solution is incorporating data from large EHR and genomic databases, providing more substantial evidence to accelerate drug development [42].

Herein, we can expect MR approaches and PGx research can go side by side in leveraging big datasets and contributing to drug development. MR studies can suggest new drug targets, as illustrated earlier in *PCSK9*-Evelocumab discovery, while PGx research can justify and confirm the biological relevance of targets presented by MR. On the other hand, when PGx studies can identify genetic variants affecting druggable proteins, MR can be used to investigate the repurposing opportunities of already licensed drugs to novel indications [15].

7. Conclusion

Undoubtedly, MR has revolutionized the epidemiological field and added a powerful tool to epidemiological studies, eventually contributing to clinical decision-making. MR lies in the interface between RCTs and observational studies, surpassing the latter by their less susceptibility to confounding and reverse causality while inferior to RCTs in their limited ability to generate guidelines [2]. Nevertheless, MR can provide unmatched value when RCTs are deficient in data about vulnerable or under-represented populations.

PGx associations provide libraries of validated genetic variants that MR can use. The power and utility of MR findings are amplified when the used IVs are of biological relevance (such as variants resulting from PGx research), rather than being selected from variants that have solely statistical significance (which are usually the variants detected in GWAS).

In drug discovery, PGx and MR exhibit a synergistic relationship that can accelerate identifying new targets and repurposing old drugs. PGx can support drug development at multiple stages, a consensus view still underapplied.

The foremost ambition of PGx researchers is to strengthen the evidence of PGx findings to accelerate their implementation at the population level. As illustrated in the work of Bowden and colleagues, MR provided supporting evidence that enabled generalizing a PGx finding at a population level and provided an answer for a public health question; how a genetically moderated treatment (clopidogrel in their study) effect is related to a common health issue (stroke prevention) [43]. This application represents how epidemiological methods, namely MR, can contribute to validating PGx practices. Indeed, we believe that the most crucial interaction would be through MR support to PGx findings acquisition. Such interactions could not be possible without the availability of big genetic data, like the biobank data and clinical data from EHRs.

The current availability of big data should be the key motivation for interdisciplinary research. Herein, we suggest further collaborations between PGx researchers, epidemiologists experts in MR applications, and data scientists to embrace the unprecedented potential in their expertise interactions.

Funding

This work was supported by the UAE Ministry of Education grant number [21M139].

CRediT authorship contribution statement

Lubna Q. Khasawneh: Writing – original draft, Data curation. **Zeina N. Al-Mahayri:** Writing – original draft, Data curation. **Bassam R. Ali:** Conceptualization, Funding acquisition, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] S.A.G. Taliun, D.M. Evans, Ten simple rules for conducting a mendelian randomization study, PLOS Comput. Biol. 17 (2021), e1009238.
[2] N.M. Davies, et al., Reading Mendelian randomisation studies: A guide, glossary, and checklist for clinicians, BMJ 362 (2018) k601.

[3] J.T. Oates, D. Lopez, Pharmacogenetics: An important part of drug development with a focus on its application, Int. J. Biomed. Investig. (2018) 1.
[4] J.A. Kim, et al., Pharmacogenomic biomarkers in US FDA-approved drug labels (2000–2020), J. Pers. Med. 11 (2021).
[5] M. Arbitrio, et al., Pharmacogenomics biomarker discovery and validation for translation in clinical practice, Clin. Transl. Sci. 14 (2021) 113–119.
[6] R. Huddart, et al., Are randomized controlled trials necessary to establish the value of implementing pharmacogenomics in the clinic? Clin. Pharmacol. Ther. 106 (2019) 284–286.
[7] F.H. van der Baan, et al., Pharmacogenetics in randomized controlled trials: Considerations for trial design, Pharmacogenomics 12 (2011) 1485–1492.
[8] L.E. Mokry, et al., Mendelian randomisation applied to drug development in cardiovascular disease: A review, J. Med Genet 52 (2015) 71–79.
[9] J. Zheng, et al., Recent developments in mendelian randomization studies, Curr. Epidemiol. Rep. 4 (2017) 330–345.
[10] E.A.W. Slob, S. Burgess, A comparison of robust Mendelian randomization methods using summary data, Genet. Epidemiol. 44 (2020) 313–329.
[11] A.I. Christensen, et al., Alcohol Intake and Risk of Ischemic and Haemorrhagic Stroke: Results from a Mendelian Randomisation Study, J. Stroke 20 (2018) 218–227.
[12] R.D. Santos, et al., Evolocumab in pediatric heterozygous familial hypercholesterolemia, N. Engl. J. Med. 383 (2020) 1317–1327.
[13] B.A. Ference, et al., Mendelian randomization study of ACLY and cardiovascular disease, N. Engl. J. Med. 380 (2019) 1033–1042.
[14] S. Park, et al., Causal effects of atrial fibrillation on brain white and gray matter volume: a Mendelian randomization study, BMC Med. 19 (2021) 274.
[15] L. Gaziano, et al., Actionable druggable genome-wide Mendelian randomization identifies repurposing opportunities for COVID-19, Nat. Med. 27 (2021) 668–676.
[16] A.F. Schmidt, et al., PCSK9 genetic variants and risk of type 2 diabetes: A mendelian randomisation study, Lancet Diabetes Endocrinol. 5 (2017) 97–105.
[17] A.G. Lai, et al., An informatics consult approach for generating clinical evidence for treatment decisions, BMC Med. Inform. Decis. Mak. 21 (2021) 281.
[18] S. Trompet, et al., Apolipoprotein E genotype, plasma cholesterol, and cancer: A mendelian randomization study, Am. J. Epidemiol. 170 (2009) 1415–1421.
[19] Z. J, et al., Genetically elevated C-reactive protein and ischemic vascular disease, N. Engl. J. Med. (2008) 359.
[20] J.C. Cohen, et al., Sequence variations in PCSK9, low LDL, and protection against coronary heart disease, N. Engl. J. Med. 354 (2006) 1264–1272.
[21] B.A. Ference, et al., Variation in PCSK9 and HMGCR and risk of cardiovascular disease and diabetes, N. Engl. J. Med. 375 (2016) 2144–2153.
[22] M.V. Holmes, et al., Secretory phospholipase A2-IIA and cardiovascular disease: A mendelian randomization study, J. Am. Coll. Cardiol. 62 (2013) 1966–1976.
[23] S.J. Nicholls, et al., Varespladib and cardiovascular events in patients with an acute coronary syndrome: the VISTA-16 randomized clinical trial, JAMA 311 (2014) 252–262.
[24] P. Linsel-Nitschke, et al., Lifelong reduction of LDL-cholesterol related to a common variant in the LDL-receptor gene decreases the risk of coronary artery disease–a Mendelian Randomisation study, PLoS One 3 (2008), e2986.
[25] M.V. Holmes, et al., Mendelian randomization of blood lipids for coronary heart disease, Eur. Heart J. 36 (2015) 539–550.
[26] Interleukin-6 Receptor Mendelian Randomisation Analysis (IL6R MR) Consortium et al. (2012) The interleukin-6 receptor as a target for prevention of coronary heart disease: a mendelian randomisation analysis. Lancet 379, 1214–1224.
[27] R. Cacabelos, et al., Genophenotypic factors and pharmacogenomics in adverse drug reactions, Int J. Mol. Sci. 22 (2021) 13302.
[28] V.M. Walker, et al., Mendelian randomization: A novel approach for the prediction of adverse drug events and drug repurposing opportunities, Int. J. Epidemiol. 46 (2017) 2078–2089.
[29] L.A. Lotta, et al., Association between low-density lipoprotein cholesterol-lowering genetic variants and risk of Type 2 diabetes: A meta-analysis, JAMA 316 (2016) 1383–1391.
[30] D.I. Swerdlow, et al., HMG-coenzyme A reductase inhibition, type 2 diabetes, and bodyweight: Evidence from genetic analysis and randomised trials, Lancet 385 (2015) 351–361.
[31] R.A.J. Smit, et al., Statin-induced LDL cholesterol response and type 2 diabetes: A bidirectional two-sample Mendelian randomization study, Pharm. J. 20 (2020) 462–470.
[32] G. Hemani, et al., Evaluating the potential role of pleiotropy in Mendelian randomization studies, Hum. Mol. Genet 27 (2018) R195–R208.
[33] G. Davey Smith, S. Ebrahim, 'Mendelian randomization': Can genetic epidemiology contribute to understanding environmental determinants of disease?*, Int. J. Epidemiol. 32 (2003) 1–22.
[34] S.J. Lewis, Mendelian randomization as applied to coronary heart disease, including recent advances incorporating new technology, Circ. Cardiovasc Genet 3 (2010) 109–117.
[35] Z. Yang, et al., Mendelian randomization study of interleukin (IL)-1 family and lung cancer, Sci. Rep. 11 (2021) 17606.
[36] S. Burgess, et al., Beyond Mendelian randomization: how to interpret evidence of shared genetic predictors, J. Clin. Epidemiol. 69 (2016) 208–216.
[37] M.E.K. Niemi, et al., Mapping the human genetic architecture of COVID-19, Nature 600 (2021) 472–477.
[38] JR: K.J. Karczewski et al. Chapter 7: Pharmacogenomics PLOS Comput. Biol. 8 2012 e1002817.
[39] B.H. Davis, N.A. Limdi, Translational pharmacogenomics: Discovery, evidence synthesis and delivery of race-conscious medicine, Clin. Pharmacol. Ther. 110 (2021) 909–925.
[40] J.C. Denny, et al., The influence of big (clinical) data and genomics on precision medicine and drug development, Clin. Pharm. Ther. 103 (2018) 409–418.
[41] S. Weiss, et al., Case-control association studies in pharmacogenetics, Pharm. J. 1 (2001) 157–158.
[42] K. Bienfait, et al., Current challenges and opportunities for pharmacogenomics: Perspective of the Industry Pharmacogenomics Working Group (I-PWG), Hum. Genet. (2021), https://doi.org/10.1007/s00439-021-02282-3.
[43] J. Bowden, et al., The Triangulation WIthin a STudy (TWIST) framework for causal inference within pharmacogenetic research, PLOS Genet. 17 (2021), e1009783.
[44] Pilling, L.C. et al. (2021) Genetic variation in activating clopidogrel: longer-term outcomes in a large community cohort, Genetic and Genomic Medicine.
[45] E.F. Magavern, et al., The interface of therapeutics and genomics in cardiovascular medicine, Cardiovasc Drugs Ther. 35 (2021) 663–676.
[46] A. Tornio, et al., Investigating real-world clopidogrel pharmacogenetics in stroke using a bioresource linked to electronic medical records, Clin. Pharm. Ther. 103 (2018) 281–286.
[47] T. Burt, S. Dhillon, Pharmacogenomics in early-phase clinical development, Pharmacogenomics 14 (2013) 1085–1097.