A critical appraisal of pharmacogenetic inference

Running title: appraising pharmacogenetic inference

Roelof AJ Smit, MD[1,2]; Raymond Noordam, PhD[2]; Saskia le Cessie, PhD[3,4];

Stella Trompet, PhD[1,2]; J. Wouter Jukema, MD, PhD[1,5]

[1]Department of Cardiology; [2]Section of Gerontology and Geriatrics, Department of Internal Medicine; [3]Department of Clinical Epidemiology; [4]Department of Medical Statistics and Bioinformatics; [5]Einthoven Laboratory for Experimental Vascular Medicine - Leiden University Medical Center, Leiden, The Netherlands

Address for Correspondence:
Roelof AJ Smit
Department of Cardiology, C5-R
Leiden University Medical Center
Albinusdreef 2
PO Box 9600, 2300 RC
Leiden, the Netherlands
Tel: +31(0)71-52-66652
Fax: +31(0)71-52-66912
E-mail: r.a.j.smit@lumc.nl

Abstract

In essence, pharmacogenetic research is aimed at discovering variants of importance to gene-treatment interaction. However, epidemiological studies are rarely set up with this goal in mind. It is therefore of great importance that researchers clearly communicate which assumptions they have had to make, and which inherent limitations apply to the interpretation of their results. This review discusses considerations of, and the underlying assumptions for, utilizing different response phenotypes and study designs popular in pharmacogenetic research to infer gene-treatment interaction effects, with a special focus on those dealing with of clinical effects of drug treatment.

Introduction

Pharmacogenetics can be thought of as a classic example of gene-environment interaction. Namely, in the search for genetic variation which can explain inter-individual drug response variability, researchers typically aim to answer the question whether a treatment effect differs between subjects with different genotypes. In other words, whether an inherited genetic variant acts as an effect measure modifier for a certain (drug) treatment.

Although the term pharmacogenetics was coined halfway through the 20th century by Fredrich Vogel (1), widespread interest into the field truly emerged with the completion of the Human Genome Project (2) (**Figure 1**). There now exist large publically available web resources and pharmacogenetic databases, made possible by methodological advances in sequencing technology and the emergence of genome-wide testing strategies (3, 4). Regrettably, contemporary pharmacogenetic research often depends on the type of study data readily available, as most epidemiological studies are not developed with pre-specified pharmacogenetic research questions in mind. Therefore, a heterogeneous body of literature exists. Collective interpretation can be difficult, as limitations and assumptions inherent to different epidemiological study designs must be recognised. Unfortunately, there also exist notable examples in the literature where authors overextend the scope and significance of their findings.

Here, we discuss considerations relating to different response phenotypes and study designs typically found throughout the pharmacogenetic literature. Though many of the considerations and pitfalls described in this paper will also apply to other types of pharmacogenetic investigations (e.g. those focussing on ADME properties), we will especially focus on studies dealing with clinical effects of drug treatment, an area where we feel invalid inference is more prevalent or at least more visible. We will clarify which conclusions may be drawn and which

limitations naturally follow from which methodological approach. Where applicable we provide illustrative examples from the field of statin pharmacogenetics, in which a diverse range of phenotypes and study designs have been combined and investigated (5). Here, we will focus specifically on investigations into the intended effects of cholesterol reduction, on the prevention of vascular events, or on the unintended occurrence of myopathy-related complaints after starting statin therapy.

Response phenotypes

Except for sharply defined clinical outcomes such as mortality, effects of treatment can often be visualised as lying on a possible spectrum of outcomes. For example, the clinical spectrum of statin-induced myopathy ranges from commonly occurring myalgia to very rare incidents of life threatening rhabdomyolysis (6). The narrow approach of dichotomization will thus lead to a loss of information and possibly reduced statistical power (7). This may particularly be the case for drug efficacy or toxicity phenotypes related to drug dosage. Furthermore, dichotomizing outcomes may induce unnecessary phenotypic heterogeneity between studies (complicating systemic reviews and meta-analyses), and might conceal possible non-linearity in the associations under investigation. Therefore, continuously distributed outcome-traits are often preferable when available. However, these outcomes come with their own challenges (e.g. non-normal distributions), and may hinder translating the results to clinically meaningful findings. For example, prior knowledge of clear clinical bimodality (e.g. disease remission) may guide researchers in choosing a response phenotype which most closely aligns with the biology of interest. In addition, dichotomous outcomes more often allow for simple visual presentation of results and categorization may mitigate the effects of including significant outliers in your analysis.

Most pharmacogenetic investigations of interest are inherently longitudinal in nature, as one wishes to measure a phenotype just before and then after a drug treatment has started. This goal corresponds to a criterion essential to causal inference, namely temporality: that exposure preceded the outcome (i.e. onset of disease or change over time in a trait) (8). Even for binary outcomes (e.g. clinical or adverse events) it will be essential to compare incidence between drug exposure categories, including the absence of drug exposure. Whenever possible, incorporating both on- and off-treatment observations into the data analysis is therefore considered superior to solely basing conclusions on data from one or more observations made on-treatment. There exist additional reasons why utilizing repeated measurements is often preferable for quantitative traits. Firstly, a single measurement is merely a snapshot of the underlying response-curve, not representative of the true response characteristics over the whole treatment phase, which is likely to differ per individual (9). Secondly, methods that do involve baseline values can eliminate much of the between-subject variability from the treatment comparison, and are therefore typically more powerful. Thirdly, limiting the analysis to a single on-treatment value ignores possible baseline imbalances between the groups, which are likely to occur in non-randomised studies. Taking these into account may help to control for confounding by (contra)indication and in distinguishing genetic effects on the response phenotype from those on off-treatment levels. Finally, having both on- and off-treatment measurements allows for the calculation of change over time, which is easy to communicate to a broad non-statistical audience.

A further consideration is the selection of a valid time interval to assess treatment response, which should be based on clinical experience. For example, a steady-state in low-density lipoprotein cholesterol (LDL-C) may be expected 4-6 weeks after start of statin treatment (10). However, when one is interested in onset of myopathy symptoms a longer period should be

considered, e.g. the mean duration of statin therapy before onset of symptoms was 6.3 months (range 0.25-48.0) in a retrospective study of 45 patients (11).

For adverse drug reactions, response phenotypes suitable for pharmacogenetic research will generally be those which appear to be strongly tied to the drug exposure. This will often depend on baseline disease incidence, whether relative effect sizes observed in large-scale studies are of apparent clinical importance, but also whether sufficient evidence supports a causal link between the drug exposure and the adverse event. Additional practical considerations such as data availability may guide or limit researchers in their investigations. For example, while it has been reliably shown that new-onset diabetes mellitus may be caused by statin therapy (12), repeated glucose measurements have historically not been assessed within statin trials. This likely explains why statin-induced glucose changes have not been examined in the pharmacogenetic setting to date.

Defining treatment effect

The observed average treatment response in a study does not always reflect the benefit of the treatment *per se*, as the context wherein this observation is made is of great importance (**Figure 2**). This is because an individuals' treatment response, defined here as the clinical outcome after starting the treatment, is not just a combination of the drug effect (i.e. the underlying (un)measured physiochemical response) and the natural course of the disease, but may also reflect secondary effects of initiating drug treatment (13, 14). Examples include placebo effects, the possibility that the individual may have been motivated to concurrently alter lifestyle habits of prognostic significance to the outcome of interest, or that the researcher or study participant may (un)knowingly influence the measurement of the endpoint if he/she is aware of the purpose

of the study (i.e. observer bias) (14). The latter issue is more likely to occur with subjective outcomes, but may be avoided through blinding both researcher and study participants.

A serious problem in non-randomized studies is the issue of confounding by (contra)indication. In routine healthcare the decision to initiate or refrain from drug treatment is based on the prognosis of the patient. Consequently, the prognoses of treated and untreated individuals in observational studies are typically not comparable. In other words, individuals with more indications for treatment are more likely to be treated, but also more likely to have a worse outcome. If this is not taken into account through study design or statistical adjustment, straightforward inference of treatment benefits may be invalid, as it could seem that treatment actually leads to worse outcomes (15). While no statistical adjustment method can fully resolve confounding by (contra)indication in observational studies if not all confounders are known, its effects should be minimized when possible. Given that genotype is set at conception and remains fixed throughout life, confounding by (contra)indication is unlikely to bias the effect estimate of a genetic variant on the outcome of interest. However, if confounding bias is present for the association between the drug exposure and the outcome of interest, this may in select cases carry over to the assessment of interaction between the genetic variant and this drug exposure (16).

In the next sections we show that the degree to which different study designs are able to avoid or disentangle these considerations is paramount to the interpretation of results and conclusions that can be drawn, also in the field of pharmacogenetics.

Study designs

Various studies are available and appropriate to answer different types of pharmacogenetic research questions, depending on the stage of drug development. Here we focus on those suitable to evaluate the effect of genetic variation on treatment efficacy and adverse drug

reactions, questions which will typically be asked after a drug has already been approved for clinical use. In addition to post-hoc subgroup analyses within a randomised controlled trial (RCT), all traditional population-based epidemiological studies can be used in this phase. However, all study designs come with underlying assumptions and limitations, and may not be able to answer all relevant questions (**Table 1**).

Our discussion here focuses mostly on sources of bias general to all epidemiology. However, a source of confounding specific to genetic epidemiology concerns population stratification (17). If there exist subgroups of individuals within the study population which differ in terms of genotype frequency and disease risk, spurious associations may arise if this is not taken into account. Typically, this can occur when individuals from different ethnic backgrounds with limited admixture are included in the same analysis (18). However, even apparently homogenous populations may contain genetically distinct subgroups (19). As larger samples will likely be more heterogeneous, population stratification will be a larger problem here (17). This should be of particular concern to researchers involved in the field of drug-gene interaction, where large studies are typically necessary to find promising signals.

Outcome-based designs

The case-control design is perhaps the most common approach for pharmacogenetic investigations into clinical effects, often focussing on adverse drug reactions. Sampling is based on the outcome, with individuals who did (cases) develop the outcome of interest being compared to those who did not (controls), with regard to drug exposure prevalence and genotype frequencies. Case-control studies can be used to assess both main effects of the genetic variant and drug exposure on the outcome, but may also assess interaction on the additive and multiplicative scale (20) (**Table 2**).

There also exist case-control studies which solely include individuals with known drug exposure, in which the analysis is limited to comparing genotype frequency between cases and controls. For the purpose of simplicity we will assume throughout the manuscript and tables that a particular susceptibility genotype is classified as being either present or absent. If it can be assumed that genotype does not associate with the outcome of interest in the absence of drug exposure, potential differences in disease occurrence between genotype groups can be interpreted as gene-treatment interactions (21). Whether this assumption is valid is highly dependent on the outcome of interest and the observation window chosen to assess this outcome. For example, this assumption is likely to hold for LDL-C reduction after statin treatment, since genetic variants are unlikely to lead to such acute (i.e. within days/weeks) and significant LDL-C changes (~30%) in absence of the drug treatment. In contrast, a treated-only case-control study on the occurrence of coronary artery disease after statin use is likely to also turn up genetic variants affecting risk in absence of statin treatment, as the underlying atherosclerotic process has a much slower onset than statin-induced LDL-C reduction.

Major benefits of the case-control design are its cost-effectiveness compared to large cohort studies, but more importantly that it is highly suited for rare (drug) outcomes. For severe adverse drug reactions, it may sometimes even form the only realistic approach to examine genetic contributions. When the outcome of interest has a continuous distribution, sampling individuals from the extremes of the outcome distribution (e.g. comparing high- with non-responders in LDL-reduction after starting statin treatment) may greatly increase statistical power when faced with budgetary restrictions for genotyping (22). However, as shown for non-responders to statin therapy in the PROspective Study of Pravastatin in the Elderly at Risk (PROSPER) trial, issues of treatment non-adherence are especially important to consider here (23). This strategy may also be promising when rare variants are investigated, as their prevalence may be greater on the extreme ends of the outcome spectrum (24).

There are some notable challenges in performing case-control studies, the first and foremost being the selection of an appropriate control group. The control group should be representative of the source population in terms of exposure distribution and genetic ancestry (e.g. European, Asian or African ancestry), and should ideally consist of individuals who would be classified as cases if they had developed the outcome of interest. In other words, controls should meet the eligibility requirements for cases except for their outcome status (20). Preferably, a geographically defined population should be the source of sampling, so the entire at-risk population can be enumerated. For hospital- or clinic-based case-control studies it may be difficult to identify this source population, as it does not correspond to a specific geographical area. For example, trauma victims referred to the hospital could live nearby or have been flown in by helicopter. In general, the catchment area for a hospital or clinic is likely to differ for different diseases, which will need to be considered when sampling controls. Similarly, as the cases of outcome-based studies on adverse drug reactions are often identified through databases it may be difficult to recruit an appropriate control group, especially since these events are often underreported (25, 26). Case-control studies nested within an existing cohort may fare better in this regard. A further risk is that cases with short survival times may be underrepresented if collection of (genetic) data occurs sometime after the event of interest.

An alternative outcome-based design is the case-only study, wherein the analysis is restricted to cases (**Table 1**). This simple approach, which can evaluate gene-treatment interaction on the multiplicative scale, assumes that genotype and drug treatment are not correlated in the population that gave rise to the cases. Under this assumption this design increases power for the test of interaction, thereby lowering the number of cases needed to be genotyped (27). Not having or being able to find a suitable control group is another reason why this may be an attractive alternative to the conventional case-control study (28). If nested in a RCT the distributions of gene and treatment can be assumed to be independent by virtue of

randomisation, making the case-only odds ratio a valid measure of gene-treatment interaction (**Table 2**). The calculated odds ratio may however (slightly) differ between case-control and case-only studies, as case-control studies estimate different population parameters (odds-, rate-, or risk-ratio), depending on how the controls were sampled (29). An example of the case-only approach in the field of statin pharmacogenetics is that by Schiffman and colleagues, who performed a genome-wide association study on coronary heart disease risk reduction when being treated with pravastatin therapy (30). In the discovery phase they solely included coronary heart disease cases from the Cholesterol and Recurrent Events (CARE) trial and the West of Scotland Coronary Prevention Study (WOSCOPS) trial, finding that 79 common genetic variants were nominally ($P<10^{-4}$) associated with differential event reduction by the therapy. To validate these results, these variants were then genotyped in an additional placebo-controlled pravastatin trial, and in all remaining patients from CARE and WOSCOPS (with or without event) (30). This study thereby exemplified how the case-only approach could be utilized as a cost-saving measure, by first screening the genome for promising signals, before including controls.

Nesting a case-only study within a cohort study can be problematic, as it is possible that genetic factors could influence the ability to tolerate therapy. Therefore, independence between genotype and treatment may not be a valid assumption. While this could also occur within an RCT, this experimental study design is more likely to have information on, and be able to include in the analysis, enrolled individuals who did not respond or had severe side effects. It has been argued that tests of gene-treatment association in controls may indicate whether genotype and treatment are truly independent in the source population, if the outcome is sufficiently rare (31). If however the assumption of gene-treatment independence is violated and ignored, the case-only approach will provide a biased interaction effect and lead to increased false-negative results (32). Another limitation of the case-only design is that main

effects of either genetic or drug treatment on the outcome cannot be estimated, and inference is limited to examining interaction on the multiplicative scale. More generally, all outcome-based designs which cannot approximate risk ratios (rare disease assumption) or risk differences (due to knowing sampling fractions) are unable to examine interaction on the additive scale, which is often of greater public health relevance (33). Due to their observational nature, outcome-based studies are additionally highly prone to confounding, selection bias (i.e. that the association between (drug) exposure and disease differs for participants who were and were not included in the study) and information bias (i.e. systematic error in the approach adopted for measuring or collecing data from a study) (20). For the last category, especially recall bias can pose an issue, which will not apply to genotype but might to drug history.

Cohort-based designs

Cohort-based designs include the cohort and treated-only designs (**Table 1**). Typically, the rate of occurrence (or recurrence) is compared between individuals with different drug exposures levels. Increasingly, population-based cohort studies are undertaken, in which an ideally random sample or even the entirety of a defined population is included in which multiple hypotheses can be evaluated. Though these relatively expensive and time-consuming studies aim to answer the same questions of causality that outcome-based designs do, the extensive and repeated phenotyping and follow-up allows for more flexibility in investigating multiple outcomes and recent, prior and repeated drug exposure (21). In addition, studying a cohort representative of a defined population allows for the calculation of population attributable risks. While this type of study typically includes more participants than outcome-based studies, it is unlikely that a single study would be able to overcome the power and sample size issues associated with genome-wide testing. Considerations of sample size are discussed in detail in

a separate section below. As cohort-based designs do not typically allow for blinding of researchers and participants, it is very likely that observer effects will not be equal between the treatment groups. In addition, if genetic testing was not undertaken close to commencement of treatment, selection bias may occur when non-responders or those with severe side effects are absent from the population.

Of greater issue is that the assignment of drug therapy is likely to have been subject driven. This means that the prognoses of the treated and untreated subjects will generally not be alike. In addition to this previously discussed confounding by (contra)indication, the issue of regression-to-the-mean may be problematic here. This occurs because the group of subjects at the extremes of the response distribution at baseline not just consists of those who consistently have more extreme values compared to the population average, but also those who simply by chance had an extreme value at baseline. Subsequent measurements of those who fall in the second category will therefore tend to be closer to the population mean thereof. Observed phenotypic changes over time may thus (partially) represent this regression-to-the-mean, which can occur when participants and/or treatment are selected on phenotypic cut-offs at baseline. This statistical phenomenon has been demonstrated for a wide range of biological measures, including lipid levels (34). Therefore, in non-randomised studies, it should be considered to combine multiple baseline measurements to reduce measurement error when selecting subjects, or to use suitable statistical methods (35, 36).

The treated-only design essentially tries to limit the issue of confounding by contraindication whilst improving statistical efficiency (37). As the name suggests, this design limits the analysis to those exposed to the drug, thereby leaving out the subjects who might have had a pertinent contraindication to treatment. This contrasts with cohorts which do include an untreated control group, in which confounding by (contra)indication is more commonly addressed through statistical adjustment, although applying stricter inclusion criteria at

13

enrolment may also limit this issue (14). A clear benefit of the treated-only approach is that less individuals are required for the analysis, which can be highly advantageous when genotyping study participants. As noted for the treated-only outcome-based design, the central assumption for inferring gene-treatment interaction effects here is that the genetic variant is unlikely to explain change in outcome in absence of the drug exposure (21). A clear drawback are that the main effects of genetic variants on the outcome are inseparable from drug-treatment interaction effects. Observed loci may thus be associated with the natural course of the disease (37). In these cases, leveraging publically available data from genome-wide association studies (GWAS) may help to substantiate the claim of absence of a main effect of a genetic variant on the outcome of interest. This approach will however require these GWAS to have taken into account possible effects of drug treatment and to have a similar outcome definition.

Of special note, an increasing number of researchers are utilizing (singular or repeated) cross-sectional data from cohort studies to perform genome-wide gene-treatment interaction analyses for quantitative traits (38). These efforts have largely been motivated by the issue that the design of many cohorts is not ideal for measuring longitudinal drug-induced changes. Specifically, assessment may be problematic when drug exposures are rare, when large intervals of time separate repeat drug exposure assessment, and when outcome phenotypes are not collected at each study visit. Therefore, the use of repeated exposure cross-sections allows for more cohorts to contribute, noting that increases in power from including more participants has been shown to be larger than the modest increase in power from making use of repeat cross-sectional measures in the same participants (39). To date, this approach has particularly been applied to questions of gene-treatment interaction for different drug classes on electrocardiography-markers (39, 40). Similar research efforts are currently underway for the field of statin pharmacogenetics.

As study information on exposure and outcome is typically determined at the same time, or at least analysed without regard for differences in time, the temporal relationship between exposure and outcome remains unclear in these cross-sectional analyses. In fact, making a distinction between exposure and outcome will generally not be possible, unless a well-established drug response phenotype is available (20). Furthermore, aside from the issues discussed previously concerning the use of a single on-treatment measurement, care must be taken to differentiate effects from those on off-treatment values. Therefore, formal comparison with an untreated group is to be advised. Alternative explanations for detected associations between genotype and outcome may be differences in number and duration of previous treatment(s) and differences in severity of disease. Using data from established cohorts may greatly facilitate the execution of these investigations. Nonetheless, due to their inherent limitations, cross-sectional studies are most suitable as hypothesis-generating tools for slowly developing diseases without sharp onset times, rather than for making solid pharmacogenetic inferences of gene-treatment interaction.

Randomized Controlled Trial

While similar in design to a cohort with a control group, the key difference for the RCT is that drug treatment is randomly allocated. As this ensures that the predictors of the outcome are equally distributed between the treated and untreated group, we can assume that: "the treated, had they remained untreated, would have experienced the same average outcome as the untreated did, and vice versa" (41). In addition, this strategy enables blinding of researcher and participant, which aims to prevent subsequent differential co-interventions or biased assessment of outcomes (42). As previously noted, if the trial is of adequate size the distributions of genotype and exposure will be independent. Due to these study characteristics,

it is possible to either avoid or account for regression-to-the-mean, confounding by (contra)indication, and selection bias. Consequently, it is possible to make more firm conclusions regarding the underlying treatment effects than is possible in non-randomised studies (**Figure 2**). While reducing the likelihood of selection bias is a major appeal of RCTs, it should be noted that genotyping in blood samples taken after study completion may still introduce this problem.

Subgroup analyses in trials have also been criticized (43), but "breaking" the randomisation will typically only occur if researchers condition on a variable that occurs after treatment, which will not apply to genotype. Though RCTs are considered the gold standard to estimate unbiased drug-SNP interaction effects, a variety of reasons exist which explain why researchers may prefer observational study settings instead. Trials will typically have included a select number of participants, thus leading to reduced statistical power compared to large observational cohorts. In addition, the relative limited number and narrow definition of exposures and outcomes under investigation may allow for less flexibility for pharmacogenetic enquiries. For example, both drug exposures and outcomes may be more clinically meaningful when examined as classes not envisioned when designing the trial. Other considerations include concerns of generalizability due to RCTs often having strict exclusion criteria, and that the RCT approach is even less suited than the cohort-based designs to investigate rare adverse outcomes. This results from individuals with relevant co-morbid conditions or with severe side effects typically being excluded before randomisation (e.g. during a run-in phase), in addition to trials often not having adequate follow-up to investigate outcomes which can occur long after the invention (44).

An approach analogous to that of the RCT, known as Mendelian randomisation, is increasingly being used in the context of pharmacogenetics and pharmacovigilance. These investigations, in which the causal effect of an exposure on an outcome is assessed by using a genetic proxy

(e.g. one or multiple genetic variants) instead of the exposure (45), have been applied to a range of different types of questions. For example, summary level statistics from a large-scale pharmacogenetic meta-analysis of GWAS of statin-induced lipid response were recently used to demonstrate that genetic predisposition for increased LDL-C levels may decrease efficacy of statin therapy if effects on off-treatment lipid levels are taken into account (46). Mendelian randomisation might alternatively be used to predict unintended drug effects. For example, Swerdlow and colleagues used SNPs in the HMGCR (i.e. the enzyme targeted by statins) gene to demonstrate that the increase in new-onset type 2 diabetes risk is "at least partially" explained by HMGCR inhibition (47). In theory, Mendelian randomisation investigations could reveal these effects prior to drugs licensing, potentially preventing exposure of large groups of patients to unnecessary risks (48). Lastly, stratifying Mendelian randomisation analyses could provide evidence which subpopulations are likely to derive greater benefit from a drug, which could guide future RCTs (49).

Considerations of sample size

A major issue in pharmacogenetic research has been the poor reproducibility of promising signals, likely in part due to underestimation of the sample sizes necessary to examine gene-treatment interaction. It has previously been demonstrated that study sizes for investigations into interaction on the multiplicative scale should be over four times as large as those necessary to detect main effects of the same magnitude (50). Given the relatively small effect sizes involved, it should therefore not come as a surprise that necessary sample sizes can run into the tens of thousands when genome-wide strategies are considered, where one must not just account for multiple testing but also consider the necessity of replicating ones results (51). Programs for sample size and power calculations for gene-treatment interaction have also been

used to estimate sample size requirements for investigations into clinical effects of statin therapy (5). In addition to study design, researchers must consider the expected sizes of both the genetic effect and the drug response, the size of their interaction effects, allele frequencies, mode of inheritance, and the prevalence of the drug treatment and outcome. Moreover, studies are likely to genotype variants in linkage disequilibrium with the true causal variant, which will also influence sample size requirements (52).

In recent years, data from mega-biobanks have been become increasingly available, which will provide unprecedented possibilities for pharmacogenetic enquiries. It should however be noted that participation rates have been relatively low, which will pose unique challenges when interpreting results. For example, only 5.2% of the 9.2 million individuals invited to enter the population-based UK Biobank actually participated in the baseline assessment (53). Similarly, in mid-2015 the Million Veterans Program estimated their response rate at 13.2% of the first 3 million invited individuals (54). In addition, it is highly questionable whether signals which can only be detected under these increased sample sizes will actually translate into clinically meaningful results.

Further considerations must be made when multiple study designs are incorporated in the same analysis via a meta-analytic approach. In the next section we will examine some of these considerations, taking the largest pharmacogenetic meta-analysis of genome-wide association studies of statin-induced LDL-C changes as an example (55).

Genomic Investigation of Statin Therapy (GIST) consortium

A major limitation of previously performed individual pharmacogenetic studies of statins was the lack of statistical power to detect small pharmacogenetics effects. To overcome this problem, a large meta-analysis of all available data on statin response was initiated, in which

the investigators aimed to combine results from statin trials and large-scale cohorts. For their meta-analysis on differential response in LDL-C to statin therapy, the GIST consortium included 6 statin-trials (n=8,421) and 10 observational studies (n=10,175) for the discovery stage. Thereafter, the most promising signals were validated in a further 22,318 subjects. Within this large GWAS effort, four loci were found to be associated with LDL-C lowering response to statin therapy. The most significant association was for a SNP on chromosome 6, at *LPA* (rs10455872, minor allele frequency (MAF)=0.08, beta=0.052, standard error (s.e.)=0.004, P=7.41x10^{-44}), indicating that carriers of the rs10455872 SNP respond to statins with a 5.2% smaller LDL-C lowering effect per minor allele compared with non-carriers. The second strongest was a SNP at *APOE* on chromosome 19 (rs445925, MAF=0.11, beta=-0.051, s.e.=0.005, P=8.52x10^{-29}), indicating an additional 5.1% increase per allele in LDL-C lowering effect compared to non-carriers. In addition, SNPs at two novel GWAS loci were shown to be significantly associated with statin response: *SORT1/CELSR2/PSRC1* at chromosome 1 (rs646776, MAF=0.22, beta=-0.013, s.e.=0.002, P=1.05x10^{-9}) and *SLCO1B1* at chromosome 12 (rs2900478, MAF=0.16, beta=0.016, s.e.=0.003, P=1.22x10^{-9}).

Notably, the consortium solely included statin-users, which made it possible to compare associations found in trials with those of observational studies. In addition, this approach made it possible to gather large enough numbers, given the necessity to account for multiple testing. To mimic the trial setting as close as possible, only incident statin users with a pre- and post-measurement were included from observational studies.

As discussed previously, the central assumption for inferring gene-treatment interaction effects via this treated-only approach is that genotype should be unlikely to significantly correlate with the response in absence of drug exposure. Given that the underlying disease course (i.e. LDL-C levels) can be assumed to be quite stable in absence of lipid-lowering treatment, this assumption may very well be valid. In addition, placebo- and observer-effects will likely be

near absent for statin-induced LDL-reduction, which will exist for more subjective complaints such as those seen within the field of psychiatric pharmacogenetics (56). The suitability of this approach was reinforced by the large homogeneity of estimates when RCTs and observational studies were separately considered.

A major point of discussion however surrounded the question how to account for the possible effect of genetic variants on off-treatment values, which cannot simply be accounted for by taking the (fractional) difference between on- and off-treatment levels as the outcome. In the end, the researchers solely included participants with on- and off-treatment LDL-C levels. Each study independently performed a GWAS on the difference between the natural log-transformed LDL-C levels on- and off-treatment which can be interpreted as the fraction of differential LDL-C lowering in carriers versus non-carriers of a genetic variant. These analyses were then adjusted for natural log-transformed off-treatment values to try to distinguish drug-treatment interaction effects from genetic effects on off-treatment LDL-C levels, a strategy for which there exists extensive debate, particularly for non-randomised studies (57, 58) By performing additional analyses, the researchers were however able to validate this approach. These included calculating formal gene-treatment interaction terms within a trial not involved in the first-stage meta-analysis for the genetic variants found to be genome-wide significant, but also by adjusting for the measurement error and intra-individual variation in off-treatment values in the only study which had multiple baseline measurements available (59).

The main limitation of the analysis is the large degree of clinical heterogeneity. This is evidenced not only by differences in eligibility criteria of the original studies, leading to the inclusion of different patient groups, but also by differences in statin types (n=8) and dosages. While adjustment for statin dose was achieved by dividing the dose by the statin-specific dose equivalent based on daily dosages required to achieve mean 30% LDL-C reduction, changes in dose during follow-up could not be taken into account. Nonetheless, the project remains a

clear example that if certain assumptions can be realistically met, inherent limitations to pharmacogenetic inference may be overcome.

Conclusion

Pharmacogenetic research is an expanding field, whose relevance is slowly becoming visible. While post-hoc subgroup comparisons in RCTs are still considered the gold standard in pharmacogenetic research of treatment efficacy, there exist many research questions for which RCTs cannot provide the solution. As all study designs and response phenotypes have their merits and problems, authors should be vigilant to avoid making conclusions which their methodology cannot back up. In particular, the assumptions needed to make inferences on gene-treatment interaction must be carefully considered, especially when case-only or treated-only strategies are employed. These challenges to inference remain ever relevant as new avenues of pharmacogenetic investigations emerge, including those using epigenetics or mRNA, as these studies will typically be performed in similar research settings.

Reference List

1.   Vogel F. Moderne Probleme der Humangenetik. Ergebn. Inn. Med. Kinderheilkd. . 1959: 12: 52-125.
2.   Lander ES, Linton LM, Birren B et al. Initial sequencing and analysis of the human genome. Nature 2001: 409: 860-921.
3.   Sim SC, Altman RB, Ingelman-Sundberg M. Databases in the area of pharmacogenetics. Hum Mutat 2011: 32: 526-531.
4.   Zhang G, Zhang Y, Ling Y et al. Web resources for pharmacogenomics. Genomics Proteomics Bioinformatics 2015: 13: 51-54.
5.   Leusink M, Onland-Moret NC, De Bakker PI et al. Seventeen years of statin pharmacogenetics: a systematic review. Pharmacogenomics 2016: 17: 163-180.
6.   Sathasivam S, Lecky B. Statin induced myopathy. BMJ 2008: 337.
7.   Steiner DL. Breaking up is hard to do: the heartbreak of dichotomizing continuous data. Can J Psychiatry 2002: 262-266.

8.   Hill AB. The environment and disease: association or causation? . Proc R Soc Med 1965: 58: 295-300.

9.   de Rotte MC, Luime JJ, Bulatovic M et al. Do snapshot statistics fool us in MTX pharmacogenetic studies in arthritis research? Rheumatology (Oxford) 2010: 49: 1200-1201.

10.   Lennernäs H, Fager G. Pharmacodynamics and pharmacokinetics of the HMG-CoA reductase inhibitors. Similarities and differences. Clin Pharmacokinet 1997: 32: 403-425.

11.   Hansen KE, Hildebrand JP, Ferguson EE et al. Outcomes in 45 patients with statin-associated myopathy. Arch Intern Med 2005: 165: 2671-2676.

12.   Collins R, Reith C, Emberson J et al. Interpretation of the evidence for the efficacy and safety of statin therapy. The Lancet 2016: 388: 2532-2561.

13.   Lubsen J, de Lang R. Klinisch geneesmiddelenonderzoek Utrecht: Bunge, 1987.

14.   Grobbee DE, Hoes AW. Clinical Epidemiology: Principles, Methods, and Applications for Clinical Research. Jones & Bartlett Learning, 2009.

15.   Grobbee DE, Hoes AW. Confounding and indication for treatment in evaluation of drug treatment for hypertension. BMJ 1997: 315: 1151-1154.

16.   Vanderweele TJ, Mukherjee B, Chen J. Sensitivity analysis for interactions under unmeasured confounding. Stat Med 2012: 31: 2552-2564.

17.   Marchini J, Cardon LR, Phillips MS et al. The effects of human population structure on large genetic association studies. Nat Genet 2004: 36: 512-517.

18.   Wang Y, Localio R, Rebbeck TR. Evaluating bias due to population stratification in epidemiologic studies of gene-gene or gene-environment interactions. Cancer Epidemiol Biomarkers Prev 2006: 15: 124-132.

19.   Smith GD, Lawlor DA, Timpson NJ et al. Lactase persistence-related genetic variant: population substructure and health outcomes. Eur J Hum Genet 2009: 17: 357-367.

20.   Rothman KJ. Modern Epidemiology. 1st Edition. Boston, MA: Little, Brown and Company. 1986.

21.   Little J, Sharp L, Khoury MJ et al. The epidemiologic approach to pharmacogenomics. Am J Pharmacogenomics 2005: 5: 1-20.

22.   Nebert DW. Extreme discordant phenotype methodology: an intuitive approach to clinical pharmacogenetics. Eur J Pharmacol 2000: 410: 107-120.

23.   Trompet S, Postmus I, Slagboom PE et al. Non-response to (statin) therapy: the importance of distinguishing non-responders from non-adherers in pharmacogenetic studies. Eur J Clin Pharmacol 2016: 72: 431-437.

24.   Auer PL, Lettre G. Rare variant association studies: considerations, challenges and opportunities. Genome Med 2015: 7: 16.

25.   Jorgensen AL, Williamson PR. Methodological quality of pharmacogenetic studies: issues of concern. Stat Med 2008: 27: 6547-6569.

26.   Eland IA, Belton KJ, van Grootheest AC et al. Attitudinal survey of voluntary reporting of adverse drug reactions. Br J Clin Pharmacol 1999: 48: 623-627.

27.   Piegorsch WW, Weinberg CR, Taylor JA. Non-hierarchical logistic models and case-only designs for assessing susceptibility in population-based case-control studies. Stat Med 1994: 13: 153-162.

28.   Khoury MJ, Flanders WD. Nontraditional epidemiologic approaches in the analysis of gene-environment interaction: case-control studies with no controls! Am J Epidemiol 1996: 144: 207-213.

29.   Schmidt S, Schaid DJ. Potential Misinterpretation of the Case-Only Study to Assess Gene-Environment Interaction. Am J Epidemiol 1999: 150: 878-885.

30.   Schiffman D, Trompet S, Louie JZ et al. Genome-wide study of gene variants associated with differential cardiovascular event reduction by pravastatin therapy. PLoS One 2012: 7: e38240.

31.   Dennis J, Hawken S, Krewski D et al. Bias in the case-only design applied to studies of gene-environment and gene-gene interaction: a systematic review and meta-analysis. Int J Epidemiol 2011: 40: 1329-1341.

32.   Albert PS, Ratnasinghe D, Tangrea J et al. Limitations of the case-only design for identifying gene-environment interactions. Am J Epidemiol 2001: 154: 687-693.

33.   Kalilani L, Atashili J. Measuring additive interaction using odds ratios. Epidemiol Perspect Innov 2006: 3: 5.

34.   Forrow L, Calkins DR, Allshouse K et al. Evaluating cholesterol screening. The importance of controlling for regression to the mean. Arch Intern Med 1995: 155: 2177-2184.

35.   Barnett AG, van der Pols JC, Dobson AJ. Regression to the mean: what it is and how to deal with it. Int J Epidemiol 2005: 34: 215-220.

36.   Chiolero A, Paradis G, Rich B et al. Assessing the Relationship between the Baseline Value of a Continuous Variable and Subsequent Change Over Time. Front Public Health 2013: 1: 29.

37.   Avery CL, Der JS, Whitsel EA et al. Comparison of study designs used to detect and characterize pharmacogenomic interactions in nonexperimental studies: a simulation study. Pharmacogenet Genomics 2014: 24: 146-155.

38.   Sitlani CM, Rice KM, Lumley T et al. Generalized estimating equations for genome-wide association studies using longitudinal phenotype data. Stat Med 2015: 34: 118-130.

39.   Avery CL, Sitlani CM, Arking DE et al. Drug-gene interactions and the search for missing heritability: a cross-sectional pharmacogenomics study of the QT interval. Pharmacogenomics J 2014: 14: 6-13.

40.   Noordam R, Sitlani CM, Avery CL et al. A genome-wide interaction analysis of tricyclic/tetracyclic antidepressants and RR and QT intervals: a pharmacogenomics study from the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium. J Med Genet 2017: 54: 313-323.

41.   Hernán MA, Robins JM. Causal Inference. Boca Raton: Chapman & Hall/CRC, forthcoming. 2017.

42.   Karanicolas PJ, Farrokhyar F, Bhandari M. Practical tips for surgical research: blinding: who, what, when, why, how? Can J Surg 2010: 53: 345-348.

43.   Sun X, Ioannidis JP, Agoritsas T et al. How to use a subgroup analysis: users' guide to the medical literature. JAMA 2014: 311: 405-411.

44.   Rothwell PM. External validity of randomised controlled trials: "To whom do the results of this trial apply?". The Lancet 2005: 365: 82-93.

45.   Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. Hum Mol Genet 2014: 23: R89-98.

46.   Smit RAJ, Postmus I, Trompet S et al. Rooted in risk: genetic predisposition for low-density lipoprotein cholesterol level associates with diminished low-density lipoprotein cholesterol response to statin treatment. Pharmacogenomics 2016: 17: 1621-1628.

47.   Swerdlow DI, Preiss D, Kuchenbaecker KB et al. HMG-coenzyme A reductase inhibition, type 2 diabetes, and bodyweight: evidence from genetic analysis and randomised trials. The Lancet 2015: 385: 351-361.

48.   Walker VM, Davey Smith G, Davies NM et al. Mendelian randomization: a novel approach for the prediction of adverse drug events and drug repurposing opportunities. Int J Epidemiol 2017. doi: 10.1093/ije/dyx207. [Epub ahead of print].

49.   Tardif JC, Rhéaume E, Lemieux Perreault LP et al. Pharmacogenomic determinants of the cardiovascular effects of dalcetrapib. Circ Cardiovasc Genet 2015: 8: 372-382.

50.   Smith PG, Day NE. The design of case-control studies: the influence of confounding and interaction effects. Int J Epidemiol 1984: 13: 356-365.

51.   Kraft P, Zeggini E, Ioannidis JP. Replication in genome-wide association studies. Stat Sci 2009: 24: 561-573.

52.   Bromley CM, Close S, Cohen N et al. Designing pharmacogenetic projects in industry: practical design perspectives from the Industry Pharmacogenomics Working Group. Pharmacogenomics J 2009: 9: 14-22.

53.   Fry A, Littlejohns TJ, Sudlow C et al. Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants with the General Population. Am J Epidemiol 2017. doi: 10.1093/aje/kwx246. [Epub ahead of print].

54.   Gaziano JM, Concato J, Brophy M et al. Million Veteran Program: A mega-biobank to study genetic influences on health and disease. J Clin Epidemiol 2016: 70: 214-223.

55.   Postmus I, Trompet S, Deshmukh HA et al. Pharmacogenetic meta-analysis of genome-wide association studies of LDL cholesterol response to statins. Nat Commun 2014: 5: 5068.

56.   Noordam R, Avery CL, Visser LE et al. Identifying genetic loci affecting antidepressant drug response in depression using drug–gene interaction models. Pharmacogenomics 2016: 17: 1029-1040.

57.   Lord EM. A paradox in the interpretation of group comparisons. Psychological Bulletin 1967: 68: 304-305.

58.   Glymour MM, Weuve J, Berkman LF et al. When is baseline adjustment useful in analyses of change? An example with education and cognitive change. Am J Epidemiol 2005: 162: 267-278.

59.   Deshmukh HA, Colhoun HM, Johnson T et al. Genome-wide association study of genetic determinants of LDL-c response to atorvastatin therapy: importance of Lp(a). J Lipid Res 2012: 53: 1000-1011.

Figure legends

Figure 1. Appearance of the terms pharmacogenetic(s) or pharmacogenomic(s) in PubMed-indexed publications across the past 25 years. The Human Genome Project was completed in 2003.

Figure 2. Non-randomized study on treatment response. The observed treatment response to drug X depends not just on the underlying physiochemical response and natural course of the disease process, but also on secondary effects of being allocated drug X. Moreover, confounding by (contra)indication may occur if reasons to initiate or refrain from drug treatment also associate with the outcome of interest. Pharmacogenetic research aims to answer which, if any, inherited genetic factors explain variation in the outcome of interest in the presence of a certain (drug) treatment (i.e. drug-gene interaction effects), distinguishing these effects from direct (i.e. main genetic effects) on the outcome.
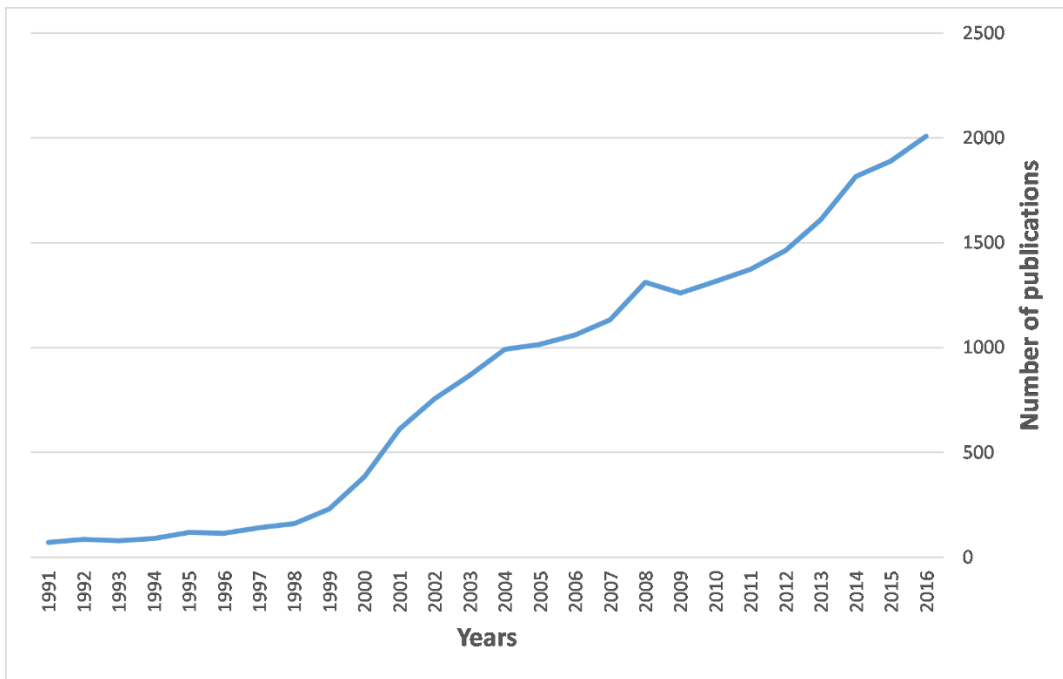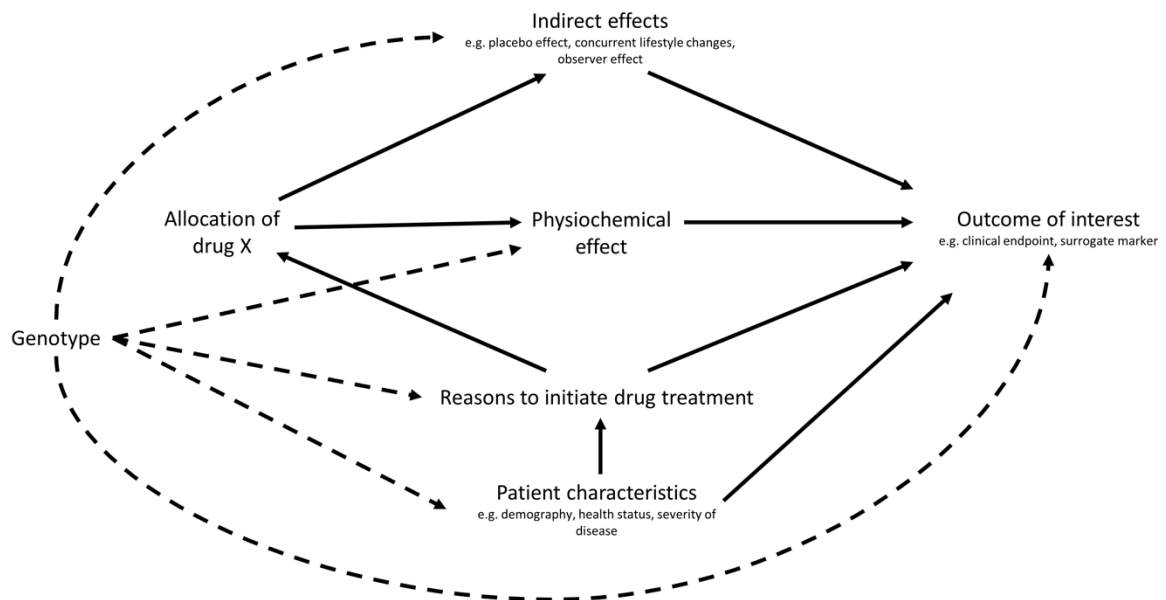
Figure 1.

Figure 2.

Table 1. Popular epidemiological study designs suitable for pharmacogenetic research questions on clinical effects of drug therapy

Design	Graphical representation	Key assumptions for gene-treatment interaction	Advantages	Limitations
Outcome-based designs				
Case-control		Valid control selection	Cost-effective; can evaluate rare events caused by rare variants; can assess both main and interaction effects	Prone to selection/information bias and confounding due to observational design
Treated-only case-control		As case-control; no association between genotype and outcome in untreated group	Genotyping untreated individuals not needed	See case-control; can only assess interaction on multiplicative scale
Case-only, nested within RCT		No association between genotype and drug exposure in source population	More efficient than case-control in evaluating interaction effects; genotyping controls not needed	See case-control; can only assess interaction on multiplicative scale; gene-treatment independence assumption unlikely to hold in non-randomised cohort
Cohort-based designs				
Cohort		-	Repeated measures; can study multiple outcomes and rare exposures; can evaluate both main and interaction effects, can assess population-attributable risk	Subject-driven assignment of treatment; resource-intensive; prone to differential loss-to-follow up (selection bias); prone to information bias and confounding; inefficient for rare outcomes
Treated-only cohort		No association between genotype and outcome in untreated group	Avoids issue of confounding by contraindication; more efficient than cohort study in evaluating interaction effects	See cohort; can only assess interaction effects; prior knowledge necessary to make key assumption for gene-treatment interaction
Trial-based design				
Subgroup analyses within RCT		Valid randomisation procedure	Random allocation of treatment assures comparability at baseline; regression-to-the-mean can be taken into account; allows for blinding	Resource-intensive; limited generalizability; inefficient for rare outcomes

RCT denotes randomised controlled trial

Table 2. Comparison of effect estimators from outcome-based study designs

Case-control setting (frequency data complete)

Drug (E)	Genotype (G)	Cases	Controls	Effect estimator
-	-	a	b	
-	+	c	d	$OR_G = b{*}c \,/\, a{*}d$
+	-	e	f	$OR_E = b{*}e \,/\, a{*}f$
+	+	g	h	$OR_{GE} = b{*}g \,/\, a{*}h$

To assess for interaction on the multiplicative scale: $OR_{GE} \,/\, (OR_G * OR_E)$

Treated-only case-control setting (subset of frequency data)

Drug (E)	Genotype (G)	Cases	Controls	Treated-only case-control $OR = f{*}g \,/\, e{*}h$
-	-	n/a	n/a	If the genetic variant G is not associated with the
-	+	n/a	n/a	outcome among untreated individuals (OR_G=1), the
+	-	e	f	treatment-only case-control OR will estimate the
+	+	g	h	assessment of interaction on the multiplicative scale
				from the case-control setting.

Case-only setting (subset of frequency data)

Drug (E)	Genotype (G)	Cases	Controls	Case-only $OR = a{*}g \,/\, c{*}e$
-	-	a	n/a	If the drug treatment E and genetic variant G are not
-	+	c	n/a	associated among controls (i.e. source population),
+	-	e	n/a	the case-only OR will estimate the assessment of
+	+	g	n/a	interaction on the multiplicative scale from the case-
				control setting.

OR denotes odds ratio. While the above table denotes genotype as the presence of absence of a certain susceptibility genotype, it will equally hold for more complex situations, including combinations of alleles at multiple loci.