

ABSTRACT

GRAHAM, HILLARY TAYLOR. Incorporating Concomitant Medications into Linear Models for Genome-Wide Association Studies. (Under the direction of Alyson Wilson.)

Given the high costs of conducting a drug-response trial, researchers are now aiming to use retrospective analyses to conduct genome-wide association studies (GWAS) in order to identify the underlying genetic contribution to drug-response variation. In order to properly perform a GWAS to investigate drug response, it is necessary to account for concomitant medications, defined as any medication taken concurrently with the primary medication being investigated, to prevent confounding of the results. We use data from the Action to Control Cardiovascular Disease (ACCORD) trial in order to implement a novel scoring procedure for incorporating concomitant medication information into a linear regression model in preparation for GWAS. In order to accomplish this, two primary medications were selected: thiazolidinediones and metformin because of the wide-spread use of these medications and large sample sizes available within the ACCORD trial. A third medication, fenofibrate, along with a known confounding medication, statin, were chosen to validate the scoring procedure. Previous studies have identified SNP rs7412 as being associated with statin response. Here we hypothesize that including the score for statin as a covariate in the GWAS model will correct for confounding of statin and yield a change in association at rs7412. The response of the confounded signal was successfully diminished from $p=5.9 \times 10^{-7}$ to $p=1.2 \times 10^{-5}$ by accounting for statin using the scoring procedure presented here. This approach allows researchers to account for concomitant medications in complex trial designs where monotherapy treatment regimens are not available.

© Copyright 2015 by Hillary Taylor Graham

All Rights Reserved

Incorporating Concomitant Medications into Linear Models
for Genome-Wide Association Studies

by
Hillary Taylor Graham

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Master of Science

Statistics

Raleigh, North Carolina

2015

APPROVED BY:

Alyson Wilson
Committee Chair

Alison Motsinger-Reif

Jung-Ying Tzeng

Daniel Rotroff

DEDICATION

To my family.

BIOGRAPHY

Hillary Graham was born in Lawton, Oklahoma in 1992. She received a Bachelors of Science in two concentrations, statistics and mathematics, from Iowa State University in 2014. In August 2014 she began the Ph.D. program in statistics at North Carolina State University.

ACKNOWLEDGMENTS

First I would like to thank my incredible committee; Dr. Alyson Wilson for providing me this opportunity and guiding me through it. Dr. Alison Motsinger-Reif and Dr. Jung Ying-Tzeng for being willing to help me in any way and answer my questions. Dr. Daniel Rotroff for putting in countless hours to provide advice, correct misconceptions, and to answer my many questions. Finally, I would like to thank my husband who has been incredibly supportive and flexible while I've been working on this project.

TABLE OF CONTENTS

LIST OF TABLES	vii
LIST OF FIGURES	viii
SECTION 1 Introduction	1
SECTION 2 Literature Review.....	6
SECTION 3 Methods.....	14
3.1 Accord Trial Data Description and Sample Selection	14
3.2 Phenotype Scores and Time-Frame Selection	16
3.3 Drug Selection	18
3.3.1 Primary Medication.....	18
3.3.2 Concomitant Medications	18
3.4 Drug Scoring.....	19
3.4.1 Primary Medication.....	19
3.4.2 Concomitant Medications	21
3.5 Non-Drug Covariate Scoring.....	22
3.6 Statistical Model Description.....	23
3.7 Validation.....	24
SECTION 4 Results.....	26
4.1 Scoring Results.....	26
4.1.1 Thiazolidinediones	26

4.1.2 Metformin	31
4.2 Validation Results	35
SECTION 5 Discussion and Conclusion	39
REFERENCES	43

LIST OF TABLES

<i>Table 1: Allocation of Subjects within the ACCORD Trial</i>	<i>14</i>
<i>Table 2: ACCORD Trial Baseline Characteristics.....</i>	<i>15</i>
<i>Table 3: Summary Statistics for change in HbA1c (%)</i>	<i>17</i>
<i>Table 4: Primary Medication Score Descriptions</i>	<i>21</i>
<i>Table 5: Annually Recorded Concomitant Medication Score Descriptions.....</i>	<i>22</i>
<i>Table 6: Distributions of Concomitant Medications within TZD Analysis.....</i>	<i>29</i>
<i>Table 7: Correlation Table for Covariates within TZD Analysis</i>	<i>29</i>
<i>Table 8: Regression Model Selected for TZD Analysis</i>	<i>30</i>
<i>Table 9: Distributions of Concomitant Medications within Metformin Analysis</i>	<i>34</i>
<i>Table 10: Correlation Table for Covariates within Metformin Analysis</i>	<i>34</i>
<i>Table 11: Regression Model Selected for Metformin Analysis.....</i>	<i>35</i>
<i>Table 12: Correlation Table for Covariates within Fenofibrate Validation</i>	<i>36</i>
<i>Table 13: Regression Models Selected for Fenofibrate Validation.....</i>	<i>37</i>

LIST OF FIGURES

<i>Figure 1: Workflow Chart</i>	3
<i>Figure 2: Variability in change in HbA1c (%)</i>	17
<i>Figure 3: Distribution of TZD Scores</i>	27
<i>Figure 4: Distribution of Concomitant Medications within TZD Analysis</i>	28
<i>Figure 5: Distribution of Metformin Scores</i>	32
<i>Figure 6: Distributions of Concomitant Medications within Metformin Analysis</i>	33
<i>Figure 7: LocusZoom plots of Chromosome 19 resulting from GWAS with and without Statin incorporated into the Model</i>	38

SECTION 1

Introduction

Interindividual variability in response to medication occurs frequently and underscores the need to develop targeted interventions. Precision medicine initiatives aim to address this issue by improving the understanding of drug mechanisms, the etiology of disease, and identifying biomarkers for drug response. However, creating new clinical trials to address these aims can be very costly. An attractive alternative is to genotype biobanked samples from previously conducted clinical trials. As the cost of genotyping continues to decrease, this approach holds potential to gain additional value from the initial investment in clinical trials.

Biobanks store and manage collections of human specimens, including but not limited to, plasma, blood, and bone marrow with the hope that samples will one day lead to breakthroughs as technology improves. In 2012, a survey of 456 biobanks in the US estimated that 59% of biobanks have been established since 2001, with each biobank containing tens to over 50 million specimens (Henderson et al. 2013). The large number of available samples and the declining costs of genotyping provide an attractive opportunity for the retrospective analysis of medical records in conjunction with biological samples (Jansen et al. 2005). Given the high costs of conducting a new drug-response trial, researchers are now aiming to use such retrospective analyses to conduct genome-wide association studies (GWAS) in order to identify the underlying genetic contribution to drug-response variation. However, clinical trials that collect biological specimens can be conducted under varying

circumstances and study designs that may not be ideal to run this drug efficacy analysis. One such example of this is the Action to Control Cardiovascular Risk in Diabetes (ACCORD) trial.

The ACCORD trial was conceived to assess the effects of controlling glycemia, blood pressure, and dyslipidemia in middle-aged and older patients with type 2 diabetes who were at high risk for cardiovascular disease (Buse 2007). The 8-year long trial was conducted at 77 clinics in the United States and Canada and followed a randomized, multicenter, double 2 x 2 factorial study design and included 10,251 patients (Buse 2007). All 10,251 ACCORD subjects participated in the glycemia trial arm, which targeted glycated hemoglobin (HbA1c), and consisted of either an intensive treatment arm targeting HbA1c less than 6% or a standard treatment arm targeting an HbA1c between 7 and 7.9%. Within each glycemia treatment arm, patients were randomized to either the blood pressure trial or the lipid trial. The blood pressure trial had an intensive arm, targeting systolic blood pressure less than 120 mm Hg, and a standard arm with a target of less than 140 mm Hg. Within the lipid trial patients were randomized to either treatments receiving fenofibrate and statin or a placebo respectively. ACCORD subjects were also asked to provide a blood sample for genotyping and 83% consented to do so.

Since ACCORD was designed to test the efficacy of different treatment strategies in individuals with type 2 diabetes, and was not designed to be a drug-response study, patients were on a variety of medications, some of which have overlapping therapeutic targets. In order to properly perform a genome-wide association study to investigate drug response, it is

necessary to account for concomitant medications, defined as any medication taken concurrently with the primary medication being investigated, to prevent confounding the results. Here we describe and implement a novel scoring procedure for incorporating concomitant medication information into a linear regression model for future genome-wide association studies using the ACCORD trial data.

Figure 1 presents a graphical representation of the scoring and model building procedures workflow. First, a primary medication is selected and scored along with a corresponding response variable (i.e. phenotype) to measure the medication's efficacy. A Wilcoxon Rank-Sum test is then utilized in order to get an idea of which concomitant medications could be affecting the response variable. Concomitant medications indicated by the Wilcoxon Rank-Sum test are then scored along with other covariates (age, smoking status, BMI, etc.). Following this, a regression model is chosen using a backwards selection criteria and the model is utilized in a GWAS.

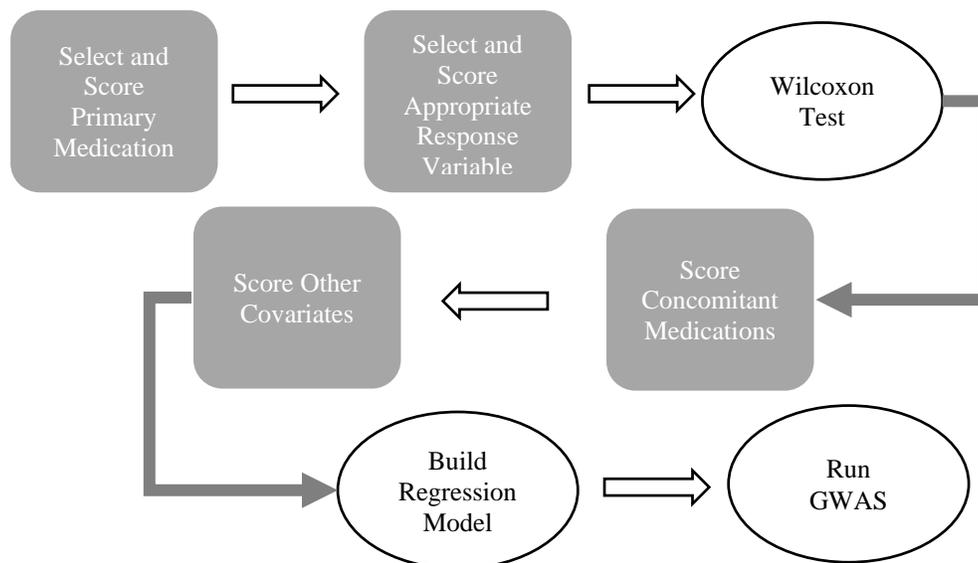


Figure 1: Workflow Chart

Here, the scoring and model building process described is applied to two primary medications, thiazolidinediones (TZDs) and metformin, which are prescribed to treat type 2 diabetes. These medications are two commonly prescribed medications to lower blood glucose and increase insulin sensitivity in diabetes patients. Both medications were prescribed in the majority of patients within the glycemia arm of the ACCORD trial. TZDs are a class of drugs that lower blood glucose by targeting the peroxisome proliferator-activated receptor gamma (PPAR γ), a member of the nuclear hormone receptor superfamily (Spiegelman 1998). It is believed that Metformin increases insulin sensitivity by targeting 5' AMP- activated protein kinase (AMPK), an enzyme that plays a diverse role in cellular energy homeostasis (Rena, Pearson, and Sakamoto 2013).

While individual patient response to these medications will vary, change in glycated hemoglobin (HbA1c) was chosen as a measure of efficacy for both TZDs and Metformin. HbA1c develops when hemoglobin binds to glucose molecules in red blood cells remains glycated for the duration of its life-cycle. Thus HbA1c is proportional to average blood glucose levels, the therapeutic target of both TZDs and metformin.

For the purpose of validating the scoring procedure, variants around the *APOE* gene will be tested for association with fenofibrate, with and without correcting for a known confounding concomitant medication, statin. Fenofibrate, prescribed to subjects within the lipid trial, lowers low-density lipoprotein (LDL) and raises high-density lipoprotein (HDL) by targeting PPAR α (peroxisome proliferator-activated receptor alpha), a nuclear receptor that plays a role in lipid metabolism (Park et al. 2006). LDL has been associated with an

increased risk of cardiovascular disease and will be the response variable for the fenofibrate validation (Cromwell et al. 2007).

Previous studies have identified the single-nucleotide polymorphism (SNP) rs7412 as being associated with statin response. Here we will use this SNP to determine if including the score for statin as a covariate in the GWAS model will correct for confounding of statin. Two regression models will be built, the first forcing statin score as a covariate and the second excluding statin as an eligible covariate.

SECTION 2

Literature Review

Properly accounting for concomitant medications is an important consideration for many drug efficacy trials. Ignoring concomitant medications can be problematic, as these secondary medications could confound the results of the primary drug of interest. Numerous methods to account for concomitant medications have been employed under differing circumstances. The intention-to-treat (ITT) concept or various implementations of dummy variables represent some approaches that are typically used to control for concomitant medications.

Information regarding the presence, dosage, and compliance of a subject's medications, concomitant or primary, is often limited or incomplete. A researcher must decide on a protocol for handling patient noncompliance, withdrawals, and missing values. Noncompliance happens when the subject is prescribed a medication but does not follow the prescriber's instructions in using it. Noncompliance ranges from forgetting to take the dose of medication one time to forgetting or refusing to take the medication numerous times which can be due to a variety of reasons including adverse side-effects. Historically, males have higher noncompliance rates than females and level of education also provides an indication of noncompliance (Khan et al. 2012). Patient withdrawal happens when a subject stops participating in the study before their role in the study is completed. This results in missing data values for this subject for the duration of the study. However, missing values can occur

in other ways as well. Data collection and data entry processes allow for human error which can often result in missing values.

ITT is the most commonly implemented method of dealing with missing outcomes, withdrawal, and patient noncompliance. ITT ignores anything that happens after subjects are randomized to treatments, including protocol deviations, subject withdrawal, and subject noncompliance (Gupta 2011). ITT is often described as “once randomized, always analyzed” (Gupta 2011).

ITT is often employed because it reflects what may happen in an actual clinical scenario (e.g. patients are not always compliant). When treatments are randomly assigned, prognostic balance is maintained and estimated treatment effect remains unbiased, arguably more so than if noncompliant and dropout patients were excluded from analysis (Gupta 2011). Another benefit of ITT is that it preserves sample size; excluding dropouts and noncompliant patients could eliminate a substantial portion of collected data and the researcher risks losing statistical power by doing so (Gupta 2011).

One of the major criticisms of ITT is that it can be too conservative in avoiding type I errors (i.e. not finding a treatment effect when one is truly present). If a patient drops out of the study before ever beginning treatment they would still be analyzed under the ITT concept. These and other similar scenarios wouldn't provide any relevant information as to the efficacy of the treatment. Thus, these cases can dilute the estimated treatment effect and result in not detecting a present effect (Gupta 2011). For these reasons ITT is often criticized for being conservative and therefore being more susceptible to type II errors (Gupta 2011).

While primarily outlined as a method for analyzing the primary medication, ITT is frequently applied to concomitant medications as information about subject compliance on secondary medications is often even more limited than that of the primary medication.

In their 2006 double blind, parallel-group atherosclerosis study, Husted et al. closely adhere to the ITT concept. Their aim was to compare the pharmacodynamics, pharmacokinetics, safety and tolerability of two antiplatelet therapies, AZD6140 and clopidogrel, in patients with atherosclerosis. Their study encompassed 201 patients, one of which dropped out before beginning treatment and 15 who withdrew after receiving at least one treatment. The patient who never received treatment was dropped from the analysis but all others were included. (Husted et al. 2006)

Certain concomitant medications were utilized as an exclusion criteria and therefore not permitted within their study. However, one concomitant medication, aspirin, was a requirement for all patients within the study and any other non-excluded concomitant medications were permitted. The most common concomitant medications within the study (> 20% of patients within any treatment group) were summarized with treatment group percentages. Husted found that these most common concomitant medications were relatively balanced across treatment groups. Because of this, concomitant medications were not accounted for in the pharmacokinetics/pharmacodynamics models used in analysis. (Husted et al. 2006)

When a study aims to measure differences across treatment groups and there are approximately equal percentages of patients taking the concomitant medication within each

treatment group, concomitant medications are often ignored. Theoretically, the effect of a secondary medication on the treatment effect will be relatively constant across treatment groups in this case. An advantage of this is that when applying a statistical model, degrees of freedom will be preserved. This is particularly relevant when the trial's sample size is small and accounting for all present concomitant medications would not be feasible. In the case of the Husted study, the researchers wanted to identify appropriate dosage amounts for further investigation in future studies with larger sample sizes (Husted et al. 2006). Including all concomitant medications in their models would have been complex and time intensive for a preliminary investigation. However, by ignoring concomitant medications, the researcher ignores important patient information that may help identify the true effect of the treatment and risks not finding an important drug-drug interactions.

Unlike Husted's study, the ACCORD trial did not employ a concomitant medication exclusion criteria and many of the medications present in the ACCORD trial have similar therapeutic targets which could confound treatment effect. Additionally, patients within each of the intensive arms of the trial (blood pressure or glycemia) were likely to be prescribed more medications than subjects in the standard arms. Because we are conducting a retrospective genetic analysis and cannot control for these factors, the ITT concept as employed by Husted et al. would not be appropriate for the ACCORD trial data. The goal of GWAS is to discover underlying genetic contribution to drug response while drug response analyses aim to find if there is differences in response. In fact many retrospective analyses using biobanked samples would have a similar problem.

In his book, *Pharmacokinetic-Pharmacodynamic Modeling and Simulation*, Bonate illustrates a few ways to use dummy variables for incorporating concomitant medications into a model (Bonate 2006). Although his book is focused on pharmacokinetic models, the methods described for incorporating concomitant medications could easily be applied to many statistical models and study designs, including the ACCORD trial. Bonate begins by suggesting an indicator for any concomitant medication. If the patient was on any concomitant medication during the study then they would receive a value of 1 for this variable and 0 if they were taking no other medications. He also mentions a variation of this approach where an ordinal variable is created. Subjects receive a 0 if they are taking no concomitant medications, a 1 if they are being minimally treated, or a 2 if the subject is being heavily treated with concomitant medications. He points out that this approach is not useful in most applications because in most studies, nearly all subjects are taking some sort of concomitant medication. Employing this method also prevents the researcher from gleaning information regarding which concomitant medications may be interacting with treatment effect and ignores all compliancy issues that may arise. (Bonate 2006)

When testing for drug-drug interactions, Bonate suggests that the researcher create one indicator per concomitant medication or group of similar concomitant medications. The variable would be coded as 0 if prior to the subject taking the medication and coded as 1 after the patient begins the concomitant therapy. Under this method, the assumption would be that the concomitant medication stays in effect throughout the end of the trial even if the patient discontinued use of the concomitant medication. Implementing a time frame around the

primary medication dosage in which the researcher believes the concomitant medication could have an interaction with the efficacy measure would be one way to get around this assumption. (Bonate 2006)

Implementing a dummy-variable system to control for concomitant medications present within a study should increase the amount of between-subject variation that is explained with the model and allow the researcher to detect drug interactions and how much a concomitant medication is influencing the treatment effect. In the case where numerous concomitant medications are present in the data, a selection criteria may need to be applied to the model to preserve degrees of freedom. However, this method still relies heavily on the assumption that patients are compliant when taking their medications and ignores any missing outcomes.

In the ACCORD trial data, systematic subject compliance information was documented for any medication the patient was taking. Ignoring compliance information under the ITT concept would be a loss of explainable variation in treatment response. While dummy variables provide an excellent way to incorporate concomitant medications, they could be modified to incorporate patient compliance data.

The dummy variable method also requires that the researcher use some sort of modeling technique. In some studies, creating a statistical model is not of interest, but controlling for concomitant medications may still be an underlying issue. In their article published in 2005, Holman and Myers describe their analysis to assess the efficacy and safety of pramipexole in fibromyalgia patients. The study included 60 fibromyalgia patients

randomized 2:1 to pramipexole or placebo respectively. An ITT analysis was used for all subjects who received at least one dose of pramipexole or placebo and had at least one followup evaluation. They implemented no concomitant medication exclusion criteria or washout periods. They did however require that all subjects taking concomitant medications strictly maintain stable dosages throughout the duration of the trial. (Holman and Myers 2005)

Prior to analyzing the primary outcome data, Holman and Myers used chi-square tests to detect significant differences between the number of subjects taking each concomitant medication within each treatment group (pramipexole or placebo). All medications but one seemed to be distributed evenly across the treatment groups. In a secondary analysis, they used analysis of covariance (ANCOVA) to detect the influence of demographic data (including concomitant medications) on the pain-score outcome. The issue of patient compliance was avoided by ensuring patients were taking steady doses of their concomitant medications throughout the duration of the trial. (Holman and Myers 2005)

While not directly related to the primary outcome variable, the chi-square tests provided useful information about subject demographics across the treatment groups. Providing these values allows the reader to examine the secondary analysis more closely. For example, in Holman and Myers study, 67% of the placebo group were using narcotic medications compared to only 44% of the pramipexole group (Holman and Myers 2005). These patient counts yielded a significant chi-square test p-value (Holman and Myers 2005). If narcotic use had an effect on pain-score outcome, then one would expect that this effect

would be stronger in the placebo group since a higher percentage of patients were using narcotics in that group.

ANCOVA is a blend between a linear regression model and analysis of variance (ANOVA). While ANCOVA has an underlying linear model, the primary objective of ANCOVA is to determine whether the dependent variable population means differ across levels of a categorical independent variable. Since ANCOVA has an underlying linear regression, it also allows for additional covariates (such as concomitant medications) to be controlled for. ANCOVA provides a way to control for concomitant medications when the primary objective is to find a difference in outcome means.

Unlike Holman and Myers' study, the ACCORD trial did not require patients to maintain steady dosages of concomitant medications throughout the trial. The trial spanned 8 years and so in many cases this would not have been possible or safe to enforce. In addition, many of the ACCORD concomitant medications were started after the trial began. When conducting a retrospective analysis with biobanked samples these and other factors are difficult to account for in subsequent analyses.

There are numerous ways to account for concomitant medications a patient may be taking when analyzing trial data. Concomitant medications can often be confounding factors in drug efficacy analyses. However, reliance on the ITT concept also has the potential to confound results by ignoring important aspects of subject compliance behavior. Here, we propose a concomitant medication scoring procedure, suitable for complex trial designs, that incorporates patient compliance in order to investigate drug response.

SECTION 3

Methods

3.1 Accord Trial Data Description

The ACCORD trial was conceived to assess the effects of controlling glycemia, blood pressure, and dyslipidemia in middle-aged and older patients with type 2 diabetes who were at high risk for cardiovascular disease (Buse 2007). All 10,251 ACCORD subjects were included in the first treatment strategy of the trial, the glycemia arm, which targets glycated hemoglobin (HbA1c). Then patients were randomized further into the lipid trial (n = 5,518 or the blood pressure trial (n = 4,733). Table 1 shows the allocation of patients within the ACCORD trial and each of the treatment group's targets.

Table 1: Allocation of Subjects within the ACCORD Trial

Glycemia Arm	Blood Pressure Trial		Lipid Trial		Total
	SBP < 120 mm Hg	SBP < 140 mm Hg	Fenofibrate	Placebo	
HbA1c < 6%	1178	1193	1383	1374	5128
HbA1c 7%-7.9%	1184	1178	1370	1391	5123
Subtotal	2362	2371	2753	2765	10251
Total	4733		5518		

Baseline characteristics for patients included in the ACCORD trial are shown below in Table 2. Overall characteristics do not differ much from characteristics within each arm of the trial. Women encompass almost 40% of the trial and the majority of trial participants were Caucasian. 58.5% of subjects were currently or had previously been a smoker. The average length of duration of diabetes was almost 11 years and the average weight of the patients was 206 lb.

Table 2: ACCORD Trial Baseline Characteristics

Baseline Characteristic	Overarching Glycemia Trial (n = 10,251)	BP Trial (n = 4,733)	Lipid Trial (n = 5,518)
Mean age (yr)	62.77	62.73	62.79
Women (%)	38.55	47.71	30.70
Race/ethnicity			
White (%)	62.36	58.76	65.46
Non-White (%)	37.64	41.24	34.54
Cigarette smoker			
Current (%)	13.94	13.23	14.55
Former (%)	44.52	42.13	46.57
Never (%)	41.54	44.64	38.88
Mean duration of diabetes (yr)	10.80	10.99	10.63
Mean weight (lb)	206.16	202.82	209.02
Mean waist circumference (in)	42.02	41.61	42.37
Mean systolic BP (mm Hg)	136.15	139.00	133.70
Mean diastolic BP (mm Hg)	74.71	75.79	73.79
Mean HbA _{1c} (%)	8.28	8.31	8.26
Mean LDL-C, mg/dL	104.72	109.60	100.53

Of the total 10,251 participants in the ACCORD trial, 83% consented to provide biological specimens to be used in future genetic analyses (Simons-Morton et al. 2014). After genotyping and quality control, 7,844 of these patients available for analysis.

3.2 Phenotype Scores and Time-Frame Selection

Phenotypes were selected to be used as a measure of efficacy (response variable) for the primary drugs of interest in our analysis. Furthermore, a standard time-frame was implemented to give the medication sufficient time to have an effect on the phenotype of interest.

Our two medications scored, TZDs and metformin, are classes of drugs that are prescribed to diabetes patients to lower blood glucose and increase insulin sensitivity as measured by glycated hemoglobin. Glycated hemoglobin (HbA1c) was chosen over fasting plasma glucose (FPG) as the measure of efficacy because HbA1c provides relatively stable samples after collection and is more convenient for the patient as FPG requires a 12 hour fast prior to collection (Nayak et al. 2012). Both of these medications were utilized in the overarching glycemia arm of the ACCORD trial.

In both analyses, the change in HbA1c was the measure of effectiveness of the medication and the response variable in all models employed. Initial phenotype values were defined as measured HbA1c at or within 30 days prior to starting the primary medication and final phenotype values were defined as the first measured Hba1c after compliance for a minimum of 90 days up to a maximum of 270 days. Since the ACCORD trial spanned 8

years and subjects did not always start taking the medication at the beginning of the trial, these time frames may fall at a different time point in the trial for each subject.

Variability in the response variable for both primary medications scored can be seen below in Figure 2. In both, the distribution of HbA1c scores seem to be approximately normal. Both distributions have a mean around -1% and a standard deviation of a little over 1%. It is evident that patients do not always have a uniform response to treatment so a model that incorporates concomitant medications may help explain some of this variability. Summary statistics for these score distributions can also be found below in Table 3.

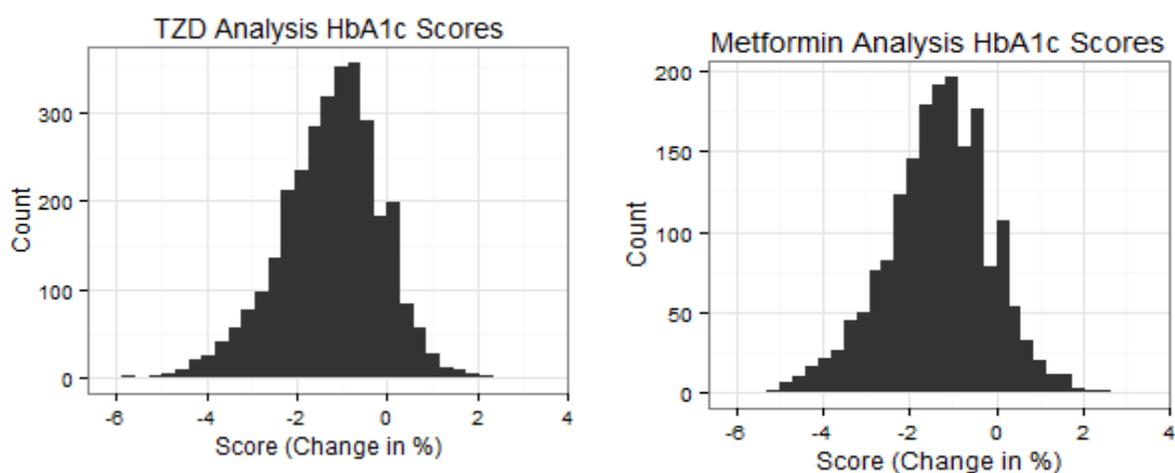


Figure 2: Variability in response variable, change in HbA1c (%). Histograms shown for TZD and Metformin patient cohorts. Distributions are approximately normal but can vary greatly from patient to patient.

Table 3: Summary Statistics for Response Variable, change in HbA1c (%)

Primary Medication	Mean	Standard Deviation	Min	Max
TZD	-1.24%	1.13%	-5.7%	3.1%
Metformin	-1.32%	1.22%	-5.7%	3.1%

3.3 Drug Selection

Drug selection was a two-part process. Primary drugs were selected to determine which genetic variants underlie variation in treatment response. Concomitant medications for each primary medication was accounted for through the scoring method. However, to avoid scoring all 72 medications present within the trial a selection criteria was implemented.

3.3.1 Primary Medications

Two medications were selected for analysis: TZDs and metformin because of the popularity of these medications and large sample sizes available within the ACCORD trial. Within the trial, 76.5% of subjects were at some point taking a TZD and 92.5% were at some point taking metformin. TZDs are a class of compounds that include rosiglitazone, pioglitazone, lobeglitazone, and troglitazone most commonly prescribed to lower blood glucose and increase insulin sensitivity.

Each patient must have maintained compliance on each of these medications for a minimum of 90 days up to a maximum of 270 days as described above. These time frames may fall at a different time point in the trial for each subject.

3.3.2 Concomitant Medications

Once the primary medication to be scored is selected, we must also account for the numerous other medications a patient was taking within the time-frame outlined above. In order to accomplish this, all 72 medications present within the Accord trial, including all blood pressure, glycemia medications, and lipid lowering medications, were tested for an association with the phenotype of interest.

The number of days a patient was on the medication was tested for association with the change in selected phenotype (HbA1c) using a Wilcoxon Rank-Sum test. The Wilcoxon Rank-Sum test, unlike parametric approaches (e.g. Student's t-test), does not require the data to be normally distributed. However, both the Student's t-test and the Wilcoxon Rank-Sum test test the same null hypothesis: that the mean of two or more populations are equal. Additionally, a false discovery rate (FDR) controlling procedure was implemented in order to account for multiple comparisons, and significantly associated concomitant medications ($q < 0.05$) were then considered eligible for selection into the model and scored according to the approach outlined in section 3.4.

Here, the Wilcoxon Rank-Sum test tests whether the number of days on the concomitant medication is associated with the change in HbA1c. Each test ignores other medications or lifestyle changes (e.g. exercise, dietary) which may have impacted the association. Thus, we expect that this initial selection approach resulted in false positives, drugs that seem to be associated with the change in HbA1c but are actually not. However, it should limit the number of false negatives, and be conservative as to not exclude any potentially confounding concomitant medications, which will be ultimately selected in the final model selection.

3.4 Drug Scoring

3.4.1 Primary Medications

In order to receive a score, all patients must have maintained full compliance (100%) in at least 80% of their recorded visits within the time-frame selected. Compliance was

recorded in the ACCORD trial as a categorical variable. A value of 1 indicated 80-100% compliance, a value of 2 indicated 1-79% compliance, 3 indicated 0% compliance, and 4 indicated the patient taking more than the prescribed dosage. A value of 4 for this scoring procedure was also considered to be full compliance (100%).

If a compliancy record was missing for a patient visit, the next non-missing compliance value was backfilled for the missing-value imputation. This differs from a standard last observation carry forward (LOCF) procedure where the last non-missing observation would be carried forward to impute missing values, whereas this approach could be described as next observation carry backward (NOCB). The NOCB approach is preferable because patient compliance is taken as a measurement for compliance since the previous visit. If patient compliance after the missing record was maintained, then it was assumed the patient was likely compliant during the missing record; whereas, a past compliancy alone does not provide information regarding the patient discontinuing the medication at a future visit.

Additionally, the patient must not have had any record of 0% compliance within the time frame to receive a score. This ensures that patients who are scored as having been on the primary medication were actually taking the medication and limits the possibility that variability in drug response is due to compliance instead of genetic variability.

A small number of patients in each analysis were also excluded due to medication records not matching across files ($n = 9$ for both TZD and metformin analyses). If a patient had records of taking the primary medication during their yearly visits but no record of the

medication could be found in their monthly visits, then the patient was excluded from further analysis.

Once ineligible patients were excluded and compliancy records were backfilled, scores were assigned as described in Table 4:

Table 4: Primary Medication Score Descriptions

Score	Description
0	The patient had no record of ever taking the primary medication of interest.
1	The patient was already on the primary medication when the trial began and stopped taking the medication before the minimum required days.
2	The patient started the primary medication at or after the trial began and stopped taking the medication before the minimum required days.
3	The patient started the primary medication at or after the trial began and was compliant for and had a measured phenotype between the minimum and maximum required days.
4	The patient was already on the primary medication when the trial began and was compliant for and had a measured phenotype between the minimum and maximum required days.

3.4.2 Concomitant Medications

Concomitant medications that were selected using the Wilcoxon Rank Sum test ($q < 0.05$) were scored using similar logic to the primary medications (Table 4). In order to receive a score, all patients must have maintained full compliance (100%) in at least 80% of their recorded visits within the time-frame selected. If a compliancy record was missing for a patient visit, the next non-missing compliance value was backfilled for the missing-value

imputation. Additionally, the patient must not have had any record of 0% compliance within the time frame to receive a score.

However, instead of the flexibility in pre and post-treatment time points afforded to the primary medication, concomitant medications were scored using the same pre and post-treatment time points defined in scoring the primary medication. Otherwise, concomitant medications with monthly records were scored using the same logic as the primary medication found in Table 4 above.

For medications with only a yearly drug records, scores were defined as described in Table 5.

Table 5: Annually Recorded Concomitant Medications Score Descriptions

Score	Description
0	The patient was not on the drug for the most recent visit prior to starting the primary medication of interest.
1	The patient was on the drug at the most recent visit prior to starting the primary medication of interest.

Within both the TZD and metformin analysis, the number of days subjects were on insulin was significantly associated with change in HbA1c. Continuous variables were not scored, but rather the average change (in mg) of total insulin per day was used in the variable selection.

3.5 Non-Drug Covariate Scoring

Medication scores were not the only covariates available for selection into the model. Other covariates included pre-treatment phenotype (HbA1c), study arm (intensive/ standard

glycemia treatment), age at the start of the trial, gender, BMI, average creatinine clearance, number of years with diabetes, number of years with dyslipidemia, smoking status, education level, glomerular filtration rate, diastolic blood pressure, systolic blood pressure, waist size, serum creatinine, network, fasting plasma glucose, alcohol consumption, and self-reported race. All covariate values were recorded as the most recent measurement prior to starting the primary medication.

Ten principal components were also utilized to avoid problems associated with population stratification. Population stratification, systematic difference in allele frequencies between subpopulations due to ancestry differences, can provide false associations if not properly accounted for (Price et al. 2010). Principal components infer genetic ancestry and can therefore help avoid the negative consequences of population stratification (Price et al. 2010).

3.6 Statistical Model Description

Once all medications and other covariates were appropriately scored, a linear regression model was constructed using all possible covariates including drug scores and additional covariates.

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_n x_n$$

Any subject with a primary medication score of 3 and no other missing covariate values were used as observational units for the regression selection. In the case that two covariates were collinear ($|r| > 0.5$), one of the covariates was dropped from the covariate

selection pool. Treatment arm indicators, principal components 1-3, to account for population stratification, and pre-treatment phenotype value were forced into the model.

Models with too many covariates are prone to overfitting, which can use too many degrees of freedom and makes results difficult to reproduce in future analyses. Thus, a backward-selection approach was used by systematically removing covariates to determine an ideal set of covariates based on the Bayesian information criterion (BIC). Backwards selection is performed by comparing BIC of the full model to the resulting BIC when removing one of the covariates. This is done for all covariates and the covariate that has the least effect on BIC is removed from selection in subsequent iterations. This is done until there is not a significant reduction in BIC when removing any of the remaining covariates.

3.7 Validation

In order to validate the scoring procedures outlined above, a third medication, fenofibrate, was chosen for analysis. Fenofibrate and statin, a known confounder, were prescribed in the lipid trial along with a double-blind placebo group; however, patients receiving placebo were excluded from the analysis. Low-density lipoprotein (LDL) was the phenotype of interest because fenofibrate is prescribed to lower LDL in diabetes patients. Final phenotype values were defined as the first measured LDL after compliance for a minimum of 90 days up to a maximum of 120 days. All other aspects of the scoring procedures are consistent with those described in Tables 4 and 5.

Two linear models were run on the fenofibrate patients. Both models used the backwards selection based on BIC as before. However, in the first model, statin scores were

added to the list of forced covariates. And in the second, statin scores were removed from the covariate selection pool.

Once the scoring and modeling procedures were complete, the DNA of patients who had been genotyped were tested using a genome-wide association study (GWAS) approach. Previous studies have identified SNP rs7412 as being associated with statin response. Here we hypothesize that including the score for statin as a covariate in the GWAS model will correct for confounding of statin and yield a change in association at rs7412.

SECTION 4

Results

4.1 Scoring Results

4.1.1 Thiazolidinediones

Within the ACCORD trial data, 2,672 of the original 10,251 subjects were not eligible to be assigned scores due to one or more of the following reasons: Patient medication information was not consistent across files (n = 9). The patient was taking a TZD, stopped for a period of time and then resumed the medication during the selected time frame (n = 439). The patient had at least one record of non-compliance (n = 379). The patient had an average compliance (after NOCB) of less than 80% within the selected time-frame (n = 809). The patient did not have an HbA1c value within the required time-frame (n = 1,832).

Of the total 10,251 subjects, 2,411 were never prescribed a TZD throughout the duration of the trial, while 57 were not on a TZD for the minimum requirement of 90 days (scores = 1 or 2). 5,111 subjects were compliant throughout the time-frame requirement; however, 2,013 of these patients were already taking a TZD at the beginning of the ACCORD trial. Thus 3,098 score 3 patients were available for genetic analysis. Figure 3 depicts the distribution of TZD scores as previously described.

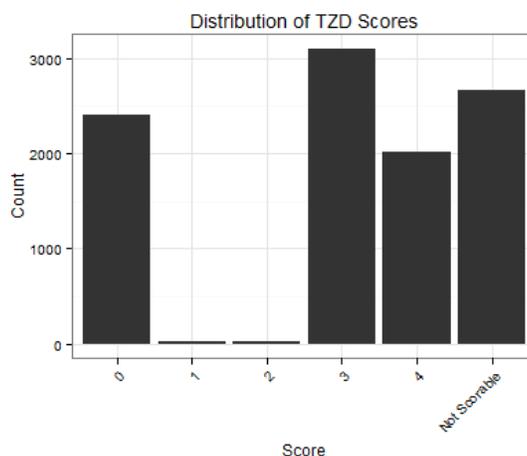


Figure 3: Distribution of TZD Scores. A large number of patients were never taking a TZD or were not scorable. Very few subjects were not compliant for long enough. 3,098 subjects had a score of 3.

The distributions of the secondary medications can be found in Figure 4 and Table 6. The population for these distributions comes from the 3,098 patients who were assigned a score of 3 for the primary medication, TZDs. Results from the initial concomitant medication screening based on Wilcoxon Rank-Sum tests are also provided in Table 6.

There were very few subjects taking meglitinide, fenofibrate, or ‘other diabetic medications’, while slightly more than half of all subjects were taking ACE inhibitors and aspirin. Metformin, glimepiride, and statin all had a large proportion of score 4s, patients who were already taking these medications at the time they started a TZD.

Since the procedure for scoring insulin yields a continuous variable, a histogram of the insulin scores is provided in the bottom right corner of Figure 4. These scores are tightly packed around 0 mg with a standard deviation of 7.75 mg indicating that few patients had a change in the amount of insulin they were taking. However, there are extreme values on either end creating long tails, indicating the majority of individuals maintained the same level

of insulin throughout the selected time-frame. The minimum insulin score is -57 mg with a maximum of 110.5 mg. The mean and median are both around 0, 0.55 mg and 0 mg respectively.

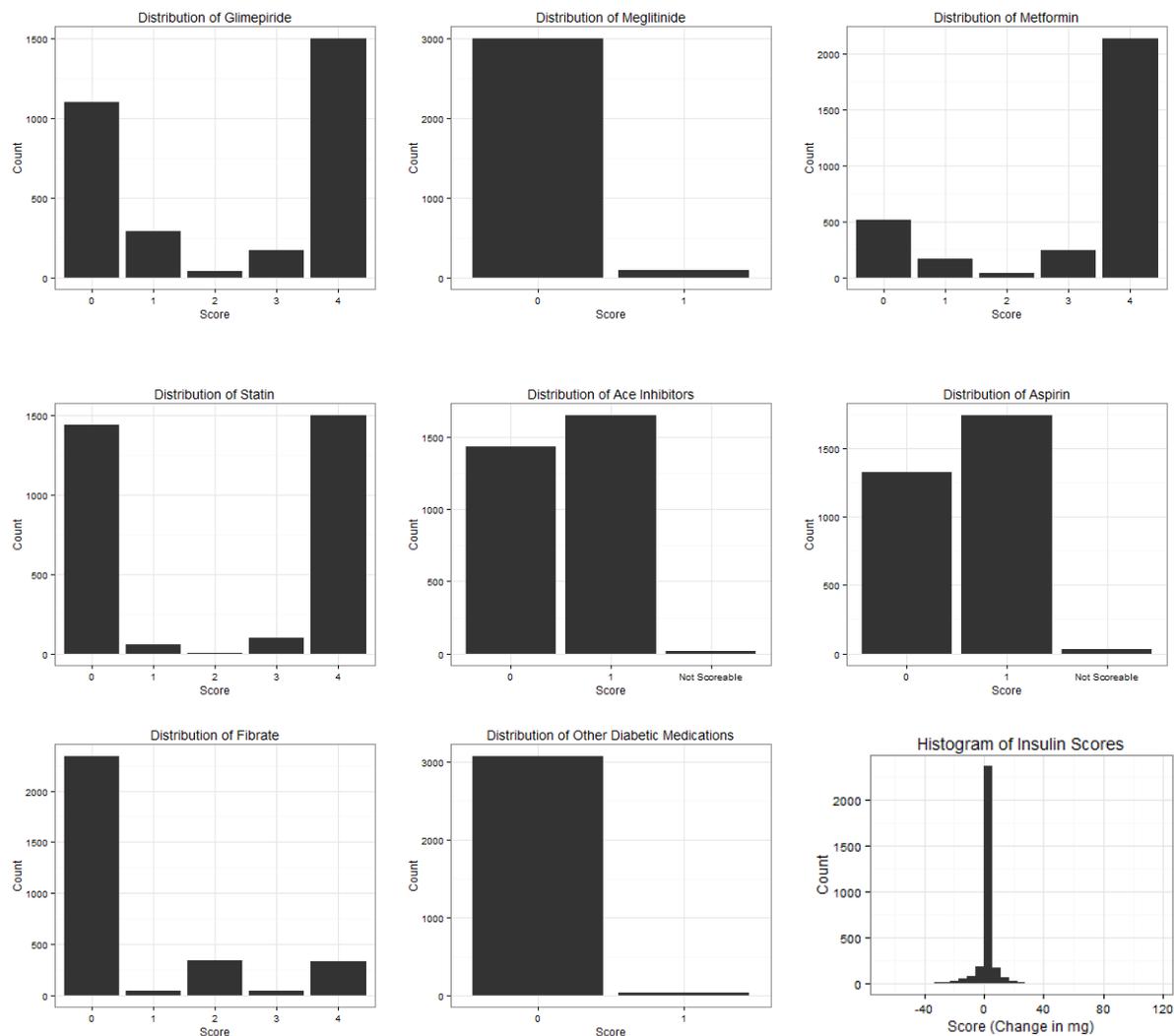


Figure 4: Distributions of Concomitant Medications within TZD Analysis. Not many subjects were taking meglitinide, fibrate, or other diabetic medications. More than half of patients were also taking aspirin or ACE inhibitors. Metformin, glimepiride, and statin have a large proportion of score 4 subjects.

Table 6: Distributions of Concomitant Medications within TZD Analysis.

Note: Cells with NA are from medications that only had yearly records and thus only have a 0 or 1 score.

Medication	q-value	p-value	Score					Not Scorable
			0	1	2	3	4	
Glimepiride	4.25*10 ⁻⁹	3.53*10 ⁻¹⁰	1099	291	37	169	1502	0
Meglitinide	9.99*10 ⁻⁹	8.66*10 ⁻¹⁰	3004	94	NA	NA	NA	0
Metformin	3.13*10 ⁻⁸	3.05*10 ⁻⁹	511	167	36	247	2137	0
Statin	2.46*10 ⁻⁶	3.04*10 ⁻⁵	1441	56	4	98	1499	0
ACE Inhibitors	0.0002	3.18*10 ⁻⁵	1433	1651	NA	NA	NA	14
Aspirin	0.0003	4.69*10 ⁻⁸	1328	1741	NA	NA	NA	29
Fibrate	0.0003	5.07*10 ⁻⁵	2341	39	340	45	333	0
Other Diabetic Medications	0.0287	0.0066	3069	29	NA	NA	NA	0

Once TZDs, the corresponding secondary medications, and all other covariates were scored, they were tested for collinearity. Table 7 shows variables that have an $|r| > 0.5$.

Variable 1 in each case was dropped for the reason also outlined in Table 7.

Table 7: Correlation Table for Covariates within TZD Analysis

Variable 1 (dropped)	Variable 2 (kept)	Correlation	Reason
Serum Creatinine	Glomerular Filtration Rate	0.57	Glomerular filtration rate incorporates serum creatinine measurements.
Waist Size (cm)	BMI	0.68	BMI incorporates the patient's size in addition to other information.
Diastolic BP	Systolic BP	0.56	Systolic blood pressure tends to be a better predictor of risk of heart attack or stroke ¹ .

Metformin	Biguanide	0.60	Metformin is in the biguanide class so should be incorporated in the biguanide score which is recorded annually and will preserve degrees of freedom in the model if selected.
Statin	Intensive BP Arm	-0.55	Including the indicator for BP arm preserves degrees of freedom.
Fasting Plasma Glucose	Pre-Treatment HbA1c	0.54	HbA1c is the outcome of interest.

¹(Mourad 2008)

After collinear variables were removed from the pool of available covariates, a linear model was selected using a backwards selection method based on BIC. Table 8 shows the final model selected. Fenofibrate was the only concomitant medication selected into the regression model.

Table 8: Regression Model selected for TZD Analysis

<i>Variable</i>	$\hat{\beta}$	$SE(\hat{\beta})$	<i>p-value</i>
Intercept	0.55	0.04	$2*10^{-16}$
Intensive Glycemia Arm	-0.78	0.03	$2*10^{-16}$
Intensive Blood Pressure	-0.04	0.04	0.40
Pre-Treatment HbA1c	-0.70	0.02	$2*10^{-16}$
Intensive Lipid Arm	0.03	0.04	0.37
Principal Component 1	-0.07	0.02	$8.65*10^{-6}$
Principal Component 2	-0.02	0.02	0.24
Principal Component 3	-0.01	0.01	0.70
Years Diabetic	0.10	0.01	$3.8*10^{-12}$
BMI	-0.09	0.02	$3.71*10^{-8}$
Fibrate Score 1	-0.15	0.09	0.09
Fibrate Score 2	-0.15	0.04	$2.6*10^{-4}$
Fibrate Score 3	0.17	0.09	0.07
Fibrate Score 4	0.08	0.05	0.10

4.1.2 Metformin

Within the Accord trial data, 1,843 of the original 10,251 subjects were not eligible to be scored due to one or more of the following reasons: Patient medication information was not consistent across files (n = 9). The patient was taking metformin, stopped for a period of time and then resumed the medication during the selected time-frame (n = 319). The patient had at least one record of non-compliance (n = 278). The patient had an average compliance (after NOCB) of less than 80% within the selected time frame (n = 919). The patient did not have an HbA1c value within the required time-frame (n = 802).

Of the total 10,251 subjects, 772 were never prescribed metformin through the duration of the trial, while 74 were not on metformin for the minimum requirement of 90 days (scores = 1 or 2). 7,562 subjects were compliant throughout the time frame requirement. However, 5,740 of these patients were already taking metformin at the beginning of the ACCORD trial, so no pre-treatment HbA1c value was available. Overall, 1,822 patients received a score of 3 and were included in the subsequent analysis. Figure 5 depicts the distribution of metformin scores as described above.

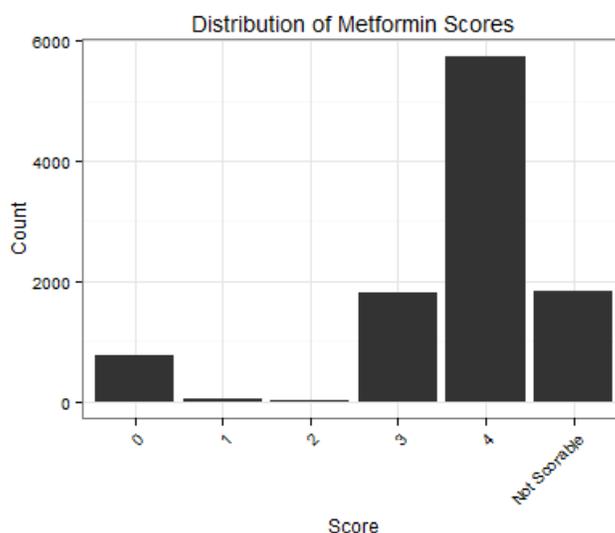


Figure 5: Distribution of Metformin Scores. A large number of patients were compliant but already taking metformin when the trial began. Very few subjects had never taken metformin or were not compliant for long enough. 1,822 subjects had a score of 3.

The distributions of the secondary medications can be found in Figure 6 and Table 9. The population for these distributions describes the 1,822 patients who were assigned a score of 3 for the primary medication, metformin. Results from the concomitant medication as determined by the Wilcoxon Rank-Sum test ($q < 0.05$) are also provided in Table 9.

For the subjects starting metformin during the selected time-frame, there were relatively few subjects prescribed meglitinide, fenofibrate, TZDs, or ‘other diabetic medications’ while slightly more than half of all subjects were taking aspirin and slightly less than half were taking ACE inhibitors. Glimepiride and statin have a large proportion of score 4s, patients who were already taking these medications at the time they started metformin.

A histogram of the insulin scores for individuals starting metformin is provided in the bottom right corner of Figure 6. These scores display little deviation from 0 mg with a standard deviation of 8.9 mg but there are extreme values on either end creating long tails.

The minimum insulin score is -67.62 mg with a maximum of 90 mg. The mean and median are both around 0 mg, -0.34 and 0 respectively.

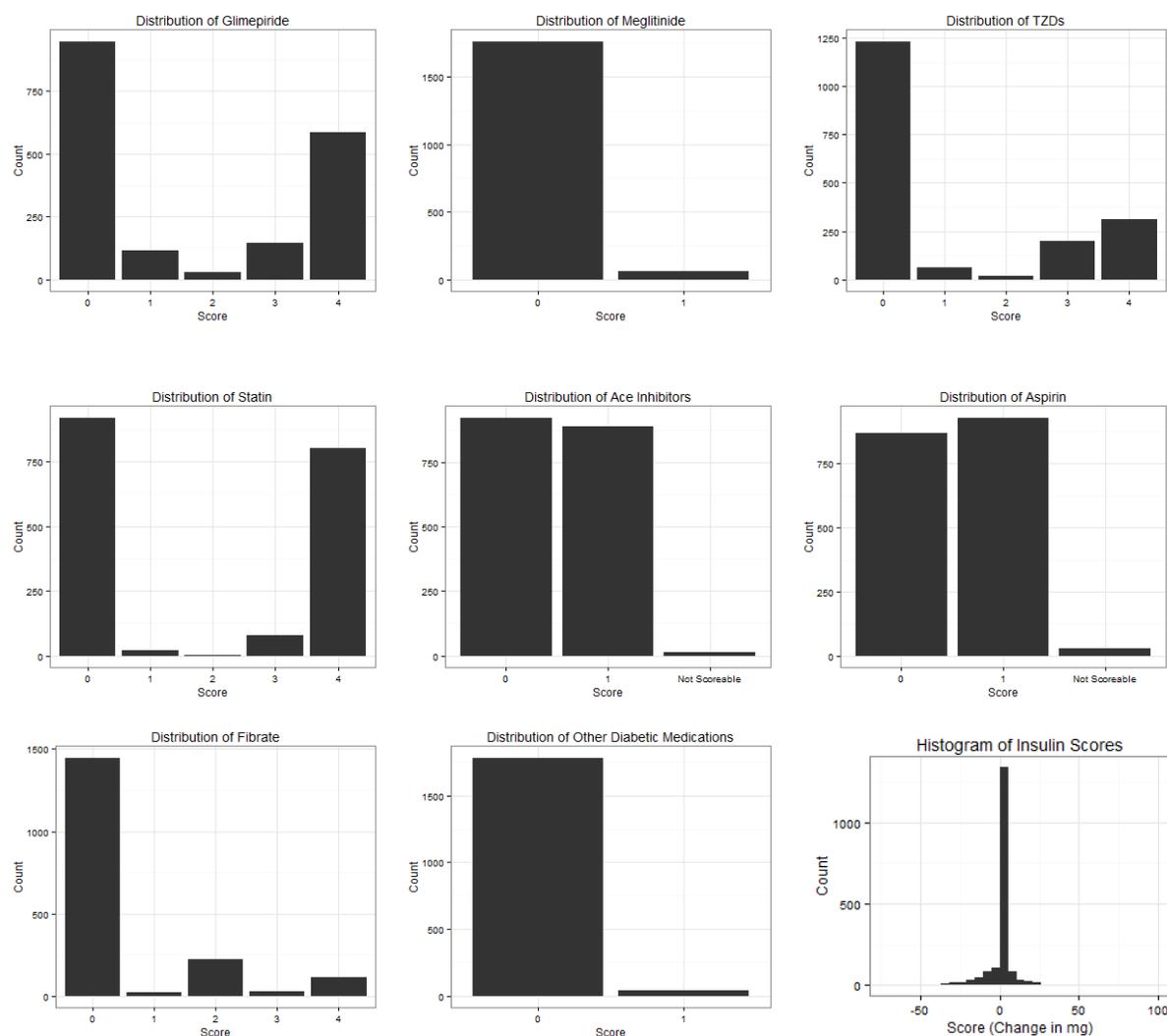


Figure 6: Distributions of Concomitant Medications within Metformin Analysis. Very few patients were taking TZDs, meglitinide, fenofibrate, or other diabetic medications. Around half of subjects were taking aspirin and ACE inhibitors. Glimepiride and statin have a large proportion of patients with a score of 4.

Table 9: Distributions of Concomitant Medications within Metformin Analysis.

Note: Cells with NA are from medications that only had yearly records and thus have a 0 or 1 score.

Medication	q-value	p-value	Score					Not Scorable
			0	1	2	3	4	
Glimepiride	4.25*10 ⁻⁹	3.53*10 ⁻¹⁰	948	114	28	147	585	0
Meglitinide	9.99*10 ⁻⁹	8.66*10 ⁻¹⁰	1760	62	NA	NA	NA	0
TZDs	2.62*10 ⁻⁸	2.45*10 ⁻⁹	1231	61	17	200	313	0
Statin	2.46*10 ⁻⁶	2.84*10 ⁻⁷	918	22	3	78	801	0
ACE Inhibitors	0.0002	3.18*10 ⁻⁵	922	887	NA	NA	NA	13
Aspirin	0.0003	4.69*10 ⁻⁵	867	927	NA	NA	NA	28
Fibrate	0.0003	5.07*10 ⁻⁵	1446	19	219	28	110	0
Other Diabetic Medications	0.0287	0.0066	1782	40	NA	NA	NA	0

Once metformin, its corresponding secondary medications, and all other covariates were scored, they were tested for collinearity. Table 10 shows variables that have an $|r| > 0.5$.

Variable 1 in each case was dropped for the reason also outlined in Table 10.

Table 10: Correlation Table for Covariates within Metformin Analysis

Variable 1 (dropped)	Variable 2 (kept)	Correlation	Reason
Serum Creatinine	Glomerular Filtration Rate	-0.77	Glomerular filtration rate incorporates serum creatinine measurements.
Waist Size (cm)	BMI	0.64	BMI incorporates the patient's size in addition to other information.
Diastolic BP	Systolic BP	0.56	Systolic blood pressure tends to be a better predictor of risk of heart attack or stroke ¹ .

Rosiglitazone	TZD	0.98	Rosiglitazone is in the TZD class so should be incorporated in the TZD score which is recorded annually and will preserve degrees of freedom in the model if selected.
Statin	Intensive BP Arm	-0.55	Including the indicator for BP arm preserves degrees of freedom.
Fasting Plasma Glucose	Pre-Treatment HbA1c	0.51	HbA1c is the outcome of interest.
Fibrate	Intensive Lipid Arm	0.81	Fibrate is the drug given in the lipid arm of the trial. Indicator for treatment arm preserves degrees of freedom.

¹(Mourad 2008)

After collinear variables were removed from the pool of available covariates, a linear model was selected using a backwards selection method based on BIC. Table 11 shows the final model selected. No concomitant medications were selected into the metformin regression model.

Table 11: Regression Model selected for Metformin Analysis

<i>Variable</i>	$\hat{\beta}$	$SE(\hat{\beta})$	<i>p-value</i>
Intercept	0.36	0.03	$2*10^{-16}$
Intensive Glycemia Arm	-0.70	0.04	$2*10^{-16}$
Intensive Blood Pressure	0.04	0.05	0.45
Pre-Treatment HbA1c	-0.69	0.02	$2*10^{-16}$
Intensive Lipid Arm	0.10	0.05	0.03
Principal Component 1	-0.08	0.02	$1.67*10^{-5}$
Principal Component 2	-0.03	0.02	0.09
Principal Component 3	0.001	0.02	0.95
Years Diabetic	0.16	0.02	$2*10^{-16}$

4.2 Validation Results

Once fenofibrate, its corresponding secondary medications, and all other covariates were scored, they were tested for collinearity. Table 12 shows variables that have an $|r| > 0.5$. Variable 1 in each case was dropped for the reason also outlined in Table 12.

Table 12: Correlation Table for Covariates within Fenofibrate Validation

Variable 1 (dropped)	Variable 2 (kept)	Correlation	Reason
Serum Creatinine	Glomerular Filtration Rate	-0.78	Glomerular filtration rate incorporates serum creatinine measurements.
Waist Size (cm)	BMI	0.64	BMI incorporates the patient's size in addition to other information.
Diastolic BP	Systolic BP	0.55	Systolic blood pressure tends to be a better predictor of risk of heart attack or stroke ¹ .
Rosiglitazone	TZD	0.66	Rosiglitazone is a TZD so should be incorporated in the TZD score which is recorded annually and will preserve degrees of freedom in the model if selected.
Metformin	Biguanide	0.75	Metformin is in the biguanide class so should be incorporated in the biguanide score which is recorded annually and will preserve degrees of freedom in the model if selected.
Glimepiride	Sulfonylurea	0.68	Glimepiride is a sulfonylurea medication so should be incorporated in the sulfonylurea score which is recorded annually and will preserve degrees of freedom in the model if selected.

¹(Mourad 2008)

After collinear variables were removed from the pool of available covariates, a linear model was selected using a backwards selection method based on BIC. Table 13 shows the final model selected when statin score is a forced covariate and when the statin scores were removed from selection. The BIC for the model with statin was 2842.36 while the model without statin had a higher BIC of 2924.77, demonstrating a better model fit while including statin, despite the incorporation of additional covariates.

Table 13: Regression Models Selected for Fenofibrate Validation

Variable	With Statin BIC = 2842.36			Without Statin BIC = 2924.77		
	$\hat{\beta}$	$SE(\hat{\beta})$	<i>p</i> -value	$\hat{\beta}$	$SE(\hat{\beta})$	<i>p</i> -value
Intercept	0.16	0.11	0.12	-0.11	0.05	0.02
Intensive Glycemia Arm	-0.09	0.04	0.02	-0.10	0.04	0.01
Principal Component 1	0.01	0.02	0.59	0.01	0.02	0.67
Principal Component 2	0.01	0.02	0.53	0.01	0.02	0.64
Principal Component 3	0.04	0.02	0.04	0.05	0.02	0.01
Pre-Treatment LDL	-0.50	0.02	$2*10^{-16}$	-0.63	0.02	$2*10^{-16}$
Years Diabetic	-0.07	0.02	$1.7*10^{-3}$	-0.07	0.02	$2.5*10^{-3}$
Years with Dyslipidemia	0.06	0.02	$4.9*10^{-3}$	0.07	0.02	$8*10^{-4}$
Statin Score 1	0.48	0.19	0.01			
Statin Score 2	-0.14	0.25	0.56			
Statin Score 3	-0.45	0.11	$3.3*10^{-5}$			
Statin Score 4	0.11	0.11	0.31			
Gender	-0.14	0.05	$1.5*10^{-3}$			
Glomerular Filtration	0.06	0.02	0.01			
Previous History of CVD				0.17	0.04	$9.4*10^{-5}$
Biguanide Score				0.14	0.05	$1.6*10^{-3}$
Principal Component 10				-0.06	0.02	0.1

LocusZoom plots of chromosome 19 from the two resulting GWAS can be found below in Figure 7. The lead SNP on the *APOE* gene (rs7412) is shown as a purple diamond. The yellow point above rs7412 is rs141622900 on the *APOC1* gene located directly to the right of *APOE* (~ 4 Kb). In the plot without statin accounted for we see a peak indicating an association near the lead SNP on the *APOE* gene ($p = 5.9*10^{-7}$). In the plot when statin was removed as a covariate, that association has been significantly reduced ($p = 1.2*10^{-5}$).

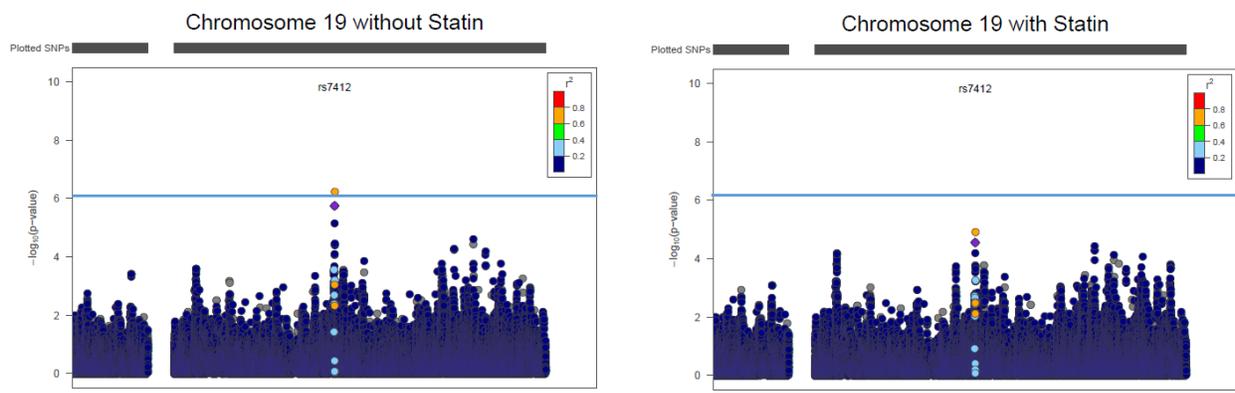


Figure 7: LocusZoom plots of Chromosome 19 resulting from the GWAS with and without statin in the model. When statin is not accounted for we see an association near the APOE gene and a lesser association when statin is accounted for.

SECTION 5

Discussion and Conclusion

Concomitant medications can be confounding variables in drug-efficacy and GWAS analyses. Medications may have overlapping mechanistic or therapeutic targets or interact in unexpected ways. However, properly accounting for potential confounding variables can be particularly challenging for data collected retrospectively and for studies not optimally designed for testing drug-response associations. Here we propose a flexible scoring scheme to control which concomitant medications are selected and how they are accounted for in the analysis model. Although this method was incorporated into a linear regression model for the purpose of conducting GWAS with the ACCORD trial data, this scoring scheme could be implemented into many statistical models.

Here we applied this scoring method to two commonly prescribed medications to treat type 2 diabetes, TZDs and metformin. Both medications were scored to identify individuals with appropriate pre- and post-treatment time points. In addition, all possible concomitant medications were tested for an association with the response variable, change in HbA1c ($q < 0.05$). Based on this, ACE inhibitors, aspirin, fenofibrate, glimepiride, insulin, meglitinide, other diabetic medications, statin were scored in both analyses. TZDs were scored as a concomitant medication in the metformin analysis and metformin was scored in the TZD analysis. All concomitant medications were scored in a similar fashion to the primary medication and eligible for inclusion in each linear model.

In the TZD model, one concomitant medication, fenofibrate, was selected for inclusion. Fenofibrate works by targeting PPAR α while TZDs target PPAR γ ; it has been shown that there is crosstalk between these receptors and both regulate aspects of lipid and glucose homeostasis (Evans, Barish, and Wang 2004). Since HbA1c, a measure of blood glucose, was the measure of response, it is not surprising that fenofibrate was incorporated into the TZD linear model as a concomitant medication.

In the metformin model, no concomitant medications were selected for inclusion. This could be due in part to the intensive glycemia arm indicator being a forced covariate. Patients within the intensive glycemia arm of the trial were more likely to be taking more medications than those in the other half of the trial. When the dummy variable for the intensive glycemia arm is removed, TZD scores were then selected into the model. The inclusion of TZD once the intensive glycemia variable was excluded suggests that the variation contributed by TZD is largely explained by the intensive glycemia arm variable. The incorporation of TZD into the metformin model is expected since they both decrease blood glucose.

A third medication, fenofibrate, was also scored along with appropriate concomitant medications and used as a validation of the approach described herein. Studies have shown that the use of statins lower LDL cholesterol, and although patient response to statins vary, part of this variability may be explained by genetics (Postmus et al. 2014). SNPs that have been shown to be associated with LDL response to statin treatment include rs2199936 (*ABCG2*), rs10455872 (*LPA*), rs7412 (*APOE*), rs445925 (*APOE*), and rs4420638 (*APOE*)

(Postmus et al. 2014). The rs7412 SNP located on the *APOE* gene is present within the ACCORD genotypes so we hypothesized that we should find an association at the *APOE* gene when we use the model without statin, while the model with statin scores accounted for should yield no such association.

Two models were used to test for associations in the fenofibrate cohort, one with statin as a forced covariate and one with statin removed as an eligible covariate. From the results of the association (Figure 7) we see an association ($p = 1.8 \times 10^{-6}$) on the *APOE* gene, but we also found a SNP (rs141622900) on the *APOC1* gene right next to *APOE* with an even stronger association ($p = 5.9 \times 10^{-7}$). Other studies have also found that SNPs located on *APOC1* may also contribute to the variation in LDL response with the use of statin (Barber et al. 2010). This demonstrates that the implemented scoring system works as expected.

The proposed scoring scheme worked well for the ACCORD trial data where a large number of study participants and detailed compliance and concomitant medication data were available. Admittedly, when retrospectively analyzing trial data these conditions will likely vary between studies, and thus this scoring scheme may not be practical for all study designs. In order to improve the sample size incorporated into the model, one could relax the time-frame requirements. Whereas, tightening the time-frame would eliminate patients without a phenotype measurement in the smaller window, but would also reduce the length of time patients were required to be compliant. Conversely, lengthening the time-frame requirement would incorporate more patients by relaxing the window for a phenotype measurement, but would require patients to have a longer record of compliancy. In future analyses it may be

informative to observe and test the differences in models selected and quality of results obtained when using these variations of the scoring scheme.

Here we have proposed a flexible scoring procedure in order to incorporate both the presence and compliance of concomitant medications in linear regression models for the purpose of finding genetic contribution to drug-response variation. This procedure could be applied to many statistical models and has been shown to decrease the confounding effect of concomitant medications.

REFERENCES

- Barber, Mathew J., Lara M. Mangravite, Craig L. Hyde, Daniel I. Chasman, Joshua D. Smith, Catherine A. McCarty, Xiaohui Li, Russell A. Wilke, Mark J. Rieder, Paul T. Williams, Paul M. Ridker, Aurobindo Chatterjee, Jerome I. Rotter, Deborah A. Nickerson, Matthew Stephens, and Ronald M. Krauss (2010), “Genome-Wide Association of Lipid-Lowering Response to Statins in Combined Study Populations,” *PLoS ONE*, 5(3), e9763.
- Bonate, Peter L. (2006), *Pharmacokinetic-Pharmacodynamic Modeling and Simulation*, Springer Science & Business Media.
- Buse, John B. (2007), “Action to Control Cardiovascular Risk in Diabetes (ACCORD) Trial: Design and Methods,” *American Journal of Cardiology*, 99(12), S21–33.
- Cromwell, William C., James D. Otvos, Michelle J. Keyes, Michael J. Pencina, Lisa Sullivan, Ramachandran S. Vasan, Peter W. F. Wilson, and Ralph B. D’Agostino (2007), “LDL Particle Number and Risk of Future Cardiovascular Disease in the Framingham Offspring Study—Implications for LDL Management,” *Journal of Clinical Lipidology*, 1(6), 583–92.
- Evans, Ronald M., Grant D. Barish, and Yong-Xu Wang (2004), “PPARs and the Complex Journey to Obesity,” *Nature Medicine*, 10(4), 355–61.
- Gupta, Sandeep K. (2011), “Intention-to-Treat Concept: A Review,” *Perspectives in Clinical Research*, 2(3), 109–12.
- Henderson, Gail E, R Jean Cadigan, Teresa P Edwards, Ian Conlon, Anders G Nelson, James P Evans, Arlene M Davis, Catherine Zimmer, and Bryan J Weiner (2013), “Characterizing Biobank Organizations in the U.S.: Results from a National Survey,” *Genome Medicine*, 5(1), 3.
- Holman, Andrew J. and Robin R. Myers (2005), “A Randomized, Double-Blind, Placebo-Controlled Trial of Pramipexole, a Dopamine Agonist, in Patients with Fibromyalgia Receiving Concomitant Medications,” *Arthritis & Rheumatism*, 52(8), 2495–2505.
- Husted, Steen, Håkan Emanuelsson, Stan Heptinstall, Per Morten Sandset, Mark Wickens, and Gary Peters (2006), “Pharmacodynamics, Pharmacokinetics, and Safety of the Oral Reversible P2Y₁₂ Antagonist AZD6140 with Aspirin in Patients with Atherosclerosis: A Double-Blind Comparison to Clopidogrel with Aspirin,” *European Heart Journal*, 27(9), 1038–47.
- Jansen, Angelique C. M., Emily S. van Aalst-Cohen, Barbara A. Hutten, Harry R. Büller, John J. P. Kastelein, and Martin H. Prins (2005), “Guidelines Were Developed for

- Data Collection from Medical Records for Use in Retrospective Analyses,” *Journal of Clinical Epidemiology*, 58(3), 269–74.
- Khan, Ataur R., Zaki N. Al-Abdul Lateef, Mohammad A. Al Aithan, Montaser A. Bu-Khamseen, Ibrahim Al Ibrahim, and Shabbir A. Khan (2012), “Factors Contributing to Non-Compliance among Diabetics Attending Primary Health Centers in the Al Hasa District of Saudi Arabia,” *Journal of Family and Community Medicine*, 19(1), 26–32.
- Mourad, Jean-Jacques (2008), “The Evolution of Systolic Blood Pressure as a Strong Predictor of Cardiovascular Risk and the Effectiveness of Fixed-Dose ARB/CCB Combinations in Lowering Levels of This Preferential Target,” *Vascular Health and Risk Management*, 4(6), 1315–25.
- Nayak, B. Shivananda, Kameel Mungrue, Shawn Seepersad, Justin Sookram, Shalina Singh, Crystal Skeete, Cadina Samaroo, Victoria Seebalack, Nirvani Siraramsingh, and Munish Sharma (2012), “Use of Hemoglobin A1c and Fasting Plasma Glucose to Diagnose Type 2 Diabetes in the Real World,” *Journal of Diabetes Science and Technology*, 6(3), 721–22.
- Park, C. W., Y. Zhang, X. Zhang, J. Wu, L. Chen, D. R. Cha, D. Su, M.-T. Hwang, X. Fan, L. Davis, G. Striker, F. Zheng, M. Breyer, and Y. Guan (2006), “PPARalpha Agonist Fenofibrate Improves Diabetic Nephropathy in Db/db Mice,” *Kidney International*, 69(9), 1511–17.
- Postmus, Iris, Stella Trompet, Harshal A. Deshmukh, Michael R. Barnes, Xiaohui Li, Helen R. Warren, Daniel I. Chasman, Kaixin Zhou, Benoit J. Arsenault, Louise A. Donnelly, Kerri L. Wiggins, Christy L. Avery, Paula Griffin, QiPing Feng, Kent D. Taylor, Guo Li, Daniel S. Evans, Albert V. Smith, Catherine E. de Keyser, Andrew D. Johnson, Anton J. M. de Craen, David J. Stott, Brendan M. Buckley, Ian Ford, Rudi G. J. Westendorp, P. Eline Slagboom, Naveed Sattar, Patricia B. Munroe, Peter Sever, Neil Poulter, Alice Stanton, Denis C. Shields, Eoin O’Brien, Sue Shaw-Hawkins, Y.-D. Ida Chen, Deborah A. Nickerson, Joshua D. Smith, Marie Pierre Dubé, S. Matthijs Boekholdt, G. Kees Hovingh, John J. P. Kastelein, Paul M. McKeigue, John Betteridge, Andrew Neil, Paul N. Durrington, Alex Doney, Fiona Carr, Andrew Morris, Mark I. McCarthy, Leif Groop, Emma Ahlqvist, Joshua C. Bis, Kenneth Rice, Nicholas L. Smith, Thomas Lumley, Eric A. Whitsel, Til Stürmer, Eric Boerwinkle, Julius S. Ngwa, Christopher J. O’Donnell, Ramachandran S. Vasan, Wei-Qi Wei, Russell A. Wilke, Ching-Ti Liu, Fangui Sun, Xiuqing Guo, Susan R Heckbert, Wendy Post, Nona Sotoodehnia, Alice M. Arnold, Jeanette M. Stafford, Jingzhong Ding, David M. Herrington, Stephen B. Kritchevsky, Gudny Eiriksdottir, Leonore J. Launer, Tamara B. Harris, Audrey Y. Chu, Franco Giulianini, Jean G. MacFadyen, Bryan J. Barratt, Fredrik Nyberg, Bruno H. Stricker, André G.

- Uitterlinden, Albert Hofman, Fernando Rivadeneira, Valur Emilsson, Oscar H. Franco, Paul M. Ridker, Vilmundur Gudnason, Yongmei Liu, Joshua C. Denny, Christie M. Ballantyne, Jerome I. Rotter, L. Adrienne Cupples, Bruce M. Psaty, Colin N. A. Palmer, Jean-Claude Tardif, Helen M. Colhoun, Graham Hitman, Ronald M. Krauss, J Wouter Jukema, and Mark J. Caulfield (2014), “Pharmacogenetic Meta-Analysis of Genome-Wide Association Studies of LDL Cholesterol Response to Statins,” *Nature Communications*, 5, <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4220464/>.
- Price, Alkes L., Noah A. Zaitlen, David Reich, and Nick Patterson (2010), “New Approaches to Population Stratification in Genome-Wide Association Studies,” *Nature reviews. Genetics*, 11(7), 459–63.
- Rena, Graham, Ewan R. Pearson, and Kei Sakamoto (2013), “Molecular Mechanism of Action of Metformin: Old or New Insights?,” *Diabetologia*, 56(9), 1898–1906.
- Simons-Morton, Denise G., Jeffrey C. Chan, Angela R. Kimel, Peter E. Linz, Cynthia L. Stowe, John Summerson, and Walter T. Ambrosius (2014), “Characteristics Associated with Informed Consent for Genetic Studies in the ACCORD Trial,” *Contemporary Clinical Trials*, 37(1), 155–64.
- Spiegelman, B. M. (1998), “PPAR-Gamma: Adipogenic Regulator and Thiazolidinedione Receptor.,” *Diabetes*, 47(4), 507–14.