Lipid Genotypes, Phenotypes, and Colorectal Polyps

Michael N. Passarelli

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2014

Reading Committee:

Polly A. Newcomb, Chair

Carolyn M. Rutter

Stephen M. Schwartz

Program Authorized to Offer Degree:

Public Health - Epidemiology

University of Washington

Abstract

Lipid Genotypes, Phenotypes, and Colorectal Polyps

Michael N. Passarelli

Chair of the Supervisory Committee:
Polly A. Newcomb, PhD, MPH
Department of Epidemiology

Studies linking cholesterol levels to colorectal neoplasia have been inconsistent. This dissertation aimed to clarify whether dyslipidemia is a risk factor for adenomatous or non-adenomatous colorectal polyps using newly-identified information on cholesterol genetics. We utilized epidemiologic data and biospecimens from a colonoscopy study conducted from 1998 to 2007 among enrollees of Group Health, a large healthcare system in Washington State. Participants were 518 non-advanced adenoma cases, 139 advanced adenoma cases, 380 non-adenomatous polyp cases, and 754 polyp-free controls. New data collected for this research included: 1) genotypes of 96 single-nucleotide polymorphisms identified from genome-wide association studies of low- and high-density lipoprotein cholesterol (LDL, HDL), triglycerides, and total cholesterol; 2) clinical cholesterol measurements from Group Health's laboratory records; and 3) information on lipid-controlling prescription drug use from Group Health's pharmacy records. Chapter 1 provides introductory results. Compared to those who reported no physician-diagnosis

of hypercholesterolemia, those who reported untreated hypercholesterolemia had increased prevalence of advanced adenomas. This association was not observed for those who reported treated hypercholesterolemia. Statin use was more likely to be associated with a decreased prevalence of advanced adenomas at the highest observed pre-colonoscopy LDL levels. Chapter 2 describes a systematic review and meta-analysis of 18 studies that assessed cholesterol from blood at the time of endoscopy, including 6,645 adenoma cases and 21,335 polyp-free controls. Individuals with colorectal adenoma were more likely than controls to have higher total cholesterol, higher triglycerides, and lower HDL. Chapter 3 describes a validation study to assess the accuracy of self-reported hypercholesterolemia, which we found to be accurate with higher specificity than sensitivity. Chapter 4 compares phenotype-polyp associations to genotype-polyp associations estimated from Mendelian randomization analyses. Given that colorectal neoplasia shares many risk factors with cardiovascular disease, including obesity, physical inactivity, and smoking, Mendelian randomization is an attractive approach to help avoid problems with confounding and reverse causation. Dyslipidemia may be a marker of the type of adiposity and dietary exposures that promote neoplasia, but there was insufficient evidence to conclude that genetic susceptibility to dyslipidemia is associated with colorectal polyps. Chapter 5 summarizes primary study limitations and recommendations for future studies.

Table of Contents

Acknowledgements

The author thanks the dissertation committee, Polly Newcomb, Michael Rosenfeld, Carolyn Rutter, and Stephen Schwartz, as well as Andrea Burnett-Hartman, Jeanne DaGloria, Karen Makar, Margaret Mandelson, Santica Marcovina, John Potter, Melissa Upton, Joseph Webster, and Lee-Ching Zhu, for their important contributions at various stages of this research.

List of Tables

List of Figures

1. Introduction and general results

Dyslipidemia and colorectal neoplasia

Dyslipidemia, usually characterized by persistently high plasma concentrations of total cholesterol (TC), triglycerides (TG), or low-density lipoprotein cholesterol (LDL), or persistently low concentrations of high-density lipoprotein cholesterol (HDL), has been found to be associated with colorectal adenomatous polyps (adenomas), established precursor lesions for colorectal cancer, but the basis for this link is not well understood (1-12). Most dyslipidemias are highly hereditary and only about 20% of circulating cholesterol is derived from diet (13). The transportation of lipoprotein macromolecules, interactions with vascular endothelium and intestinal epithelium, role in the synthesis of steroid hormones and soluble vitamins, and eventual peroxidation involve several biological pathways (14). These include cell signaling (15, 16), cell adhesion (17), growth factor regulation (18, 19), angiogenesis (20, 21), apoptosis (22, 23), and immune response (24, 25) – all of which also play important roles in carcinogenesis (26, 27). This dissertation aimed to clarify whether dyslipidemia is a risk factor for adenomatous or non-adenomatous colorectal polyps.

Using cholesterol genotypes to address problems with confounding and reverse causation

Observational studies that evaluate associations between cholesterol and chronic diseases are often plagued by confounding and reverse causation (28). Because colorectal neoplasia shares many risk factors with cardiovascular disease, including obesity, physical inactivity, and smoking behaviors, it has been challenging for epidemiological investigations to determine if the co-occurrence of dyslipidemia and precancerous colorectal lesions is indicative of shared risk factors or an etiologic link (29). One approach to help avoid problems with confounding and

reverse causation is Mendelian randomization (30). Mendelian randomization studies evaluate outcomes with respect to genetic proxies for observed phenotypes (31). While it is possible that some confounding can still occur as a consequence of population stratification and linkage disequilibrium, much of the confounding from health behaviors not directly caused by functional genetic variants is minimized by the natural randomization derived from the independent assortment of alleles during gamete formation (32). This approach has proved especially effective for studies of cholesterol and cardiovascular disease (33, 34).

More than a decade ago, genotypic associations with risk of colorectal neoplasia were noted for several important genes related to blood lipids, including the apolipoprotein genes (e.g., APOE, APOB) (35-39). These studies, however, were based on relatively small samples and only focused on single candidate genes. Today, based on large-scale genotyping studies, nearly 100 genes have been found to influence blood lipid levels. Moreover, we now know a great deal about which genes influence the different types of cholesterol and the magnitude and direction of effect for specific alleles (40).

In 2010, the Global Lipids Genetics Consortium (Teslovich, et al.) published a combined analysis of 46 GWASs involving plasma lipid measurements from over 100,000 individuals (41). Their evaluation identified 102 single nucleotide polymorphisms (SNPs) in 95 genes significantly associated with plasma concentrations of LDL, HDL, TG, or TC. To determine the association between genes known to influence blood lipid phenotypes and colorectal neoplasia, we will conduct a study of GWAS-identified SNPs that influence lipid concentrations reported by Teslovich et al. Our study will test the association between adenomas and blood concentrations of LDL, HDL, TG, and TC in our study population, and determine the association between adenomas and 102 GWAS-identified SNPs related to blood lipids. Our analysis will utilize existing epidemiologic data, medical history, and biospecimens from a completed

colonoscopy study (R01CA097325, P01CA74184) among enrollees of Group Health, a large healthcare system that serves Washington State.

Public health significance

It is estimated that up to 45% of asymptomatic individuals between the ages of 60 and 69 have at least one prevalent adenoma (42), and about 30% of adults without clinical diagnosis of cardiovascular disease have persistently high LDL, high TG, or low HDL (43). Given the rapidly increasing prevalence of chronic disease in the population, there is an urgent need to understand the biological basis by which dyslipidemia is associated with the development of precursor lesions for colorectal cancer. Cholesterol levels can be controlled through a variety of interventions, making this exposure a potentially modifiable risk factor for colorectal cancer. Cholesterol-lowering medications, particularly statins, are effective and widely used in this country (44). Studies have been inconsistent about whether statins alter the risk of developing adenoma or colorectal cancer (45-48). Our study will help inform future evaluations of statins and other cholesterol-lowering interventions for the prevention of colorectal neoplasia.

The knowledge obtained from our proposed study may help better identify those at risk of developing precursors for colorectal cancer. Endoscopy screening and treatment of dyslipidemia are two well-established clinical practices that will obviously continue, but in the future they may be considered mutually beneficial for the prevention of adenomas. Ultimately, the identification of novel genetic pathways to malignancy in the colon and rectum though lipid pathways could support the incorporation of hereditary factors of dyslipidemia into clinical guidelines for prevention of polyps, and may lead to new avenues of targeted prevention by

helping to identify which patients with abnormal cholesterol levels would benefit from increased colorectal cancer screening and surveillance.

Introduction to the study population

Details of data collection procedures will be provided in Chapters 2-4. In brief, we utilized existing epidemiologic data, medical history, and biospecimens from a completed colonoscopy-based case-control study among enrollees of Group Health, a large healthcare system in Washington State (49). New data collected for this research included genotypes obtained from blood or buccal sample of germline loci previously found to be associated with cholesterol levels from genome-wide association studies (GWAS), clinical cholesterol measurements (pre-colonoscopy) from Group Health's electronic laboratory records, and information on lipid-controlling prescription drug use (pre-colonoscopy) from Group Health's pharmacy records.

In this study, all participants received a colonoscopy for any indication. Those found to have colorectal polyps served as cases (3 case groups defined by lesion pathology were considered, see Chapter 4), and those determined to be polyp-free during colonoscopy served as controls. Eligible participants were required to have been continuously enrolled at GH for at least three years and could not have had a colonoscopy within one year of the study colonoscopy. Those with a prior diagnosis of colorectal cancer, ulcerative colitis, Crohn's disease, Lynch syndrome or familial adenomatous polyposis (FAP) were ineligible.

All participants completed a structured 40-minute telephone interview that elicited information on personal medical history, family cancer history, a brief survey of dietary practices, use of non-steroidal anti-inflammatory drugs (NSAIDs) and hormone therapy, body

size, reproductive experiences, smoking and alcohol use, physical activity and screening utilization. Nearly 75% of eligible Group Health members invited to participate completed the questionnaire.

The original data collection took place in two phases. Phase I of the study enrolled 25% of the total sample size from September 1998 to March 2003. Phase II of the study enrolled the remaining 75% of participants from December 2004 to September 2007. Protocols for eligibility were similar in both phases, and have been described in detail in several published reports from this study population (50-56). Additional details regarding the previous collection of biospecimens and pathology information are provided in Chapter 4.

New data collection for this dissertation, including genotyping and electronic record linkages, took place between November 2012 and June 2013. Details of the extraction of clinical cholesterol measurements and prescription fills of lipid-controlling drugs are included in Chapter 3. SNP-selection and genotyping methods for GWAS-identified loci known to be associated with blood levels of LDL, HDL, TG, and TC are available in Chapter 4.

Overview of the dissertation

This dissertation is divided into 3 primary chapters (Chapters 2-4). First, we review the literature on endoscopy studies that assessed blood cholesterol concentrations from a fasting blood draw at the time of endoscopy ("Chapter 2: Blood lipid concentrations and colorectal adenomatous polyps: a systematic review, meta-analysis, and meta-regression of case-control endoscopy studies, 1986-2010"). Next, we conducted a validation study to assess the accuracy of different ways that hypercholesterolemia can be self-reported ("Chapter 3: Addressing ambiguity from self-report of hypercholesterolemia not treated by lipid-controlling drugs"). The primary aims of the dissertation are addressed within the context of a Mendelian randomization study,

which evaluates colorectal polyp prevalence in relation to genetic proxies for observed cholesterol phenotypes. Mendelian randomization estimates of association are compared to estimates of association measured directly from the cholesterol phenotypes themselves ("Chapter 4: Blood lipids and colorectal polyps: testing an etiologic hypothesis using phenotypic measurements and Mendelian randomization"). A final chapter (Chapter 5) provides a summary of our main conclusions, as well as the primary limitations of this work, and a brief discussion of future directions.

Preliminary results

To conclude this chapter, we summarize some preliminary results on the association between self-report of hypercholesterolemia, assessed during the study questionnaire, and the main effect of statin use, assessed from pharmacy records, to provide a context for the primary findings of Chapters 2-4. Self-report of hypercholesterolemia is evaluated with respect to a validity analysis in Chapter 3, and would have been the only way to assess hypercholesterolemia status without the new data collection conducted specifically for this dissertation. The latter is also used to validate self-report of hypercholesterolemia requiring drug treatment in Chapter 3, and is considered as a key stratification variable in the assessment of associations between polyp prevalence and genotypes and phenotypes in Chapter 4. Details of the data collection procedures, case definitions, study exclusions, and statistical methods are described in Chapters 3 and 4.

Self-reported hypercholesterolemia and polyp prevalence

We estimated the association between colorectal polyps (non-advanced adenomas, advanced adenomas, and non-adenomatous polyps) and self-report of hypercholesterolemia ascertained during the study interview. For hypercholesterolemia status, the questionnaire asked: "Has a doctor ever told you that you have high cholesterol?" (allowable answers: yes, no, don't know). For those who answered "yes", a follow-up question asked: "Have you or are you currently taking medications for the condition?" (allowable answers: yes, no, don't know). hypercholesterolemia was defined as a categorical variable with 3 levels (the validity of which was evaluated in Chapter 3): 1) no hypercholesterolemia; 2) untreated hypercholesterolemia (received a physician-diagnosis of hypercholesterolemia, but did not take lipid-controlling drugs), and treated hypercholesterolemia (received a physician-diagnosis of hypercholesterolemia, and took lipid-controlling drugs).

Polytomous logistic regression models were used to estimate odds ratios (ORs) and 95% confidence intervals (CIs) adjusted for age at colonoscopy, sex, race, education, body mass index, nonsteroidal anti-inflammatory drug use, family history of colorectal cancer, estrogen-only use (in women), estrogen-plus-progestin use (in women), cigarette smoking, alcohol consumption, diabetes mellitus, fruit servings per day, vegetable servings per day, recreational and exercise physical activity, prior endoscopy (approximately 2 years before study colonoscopy), and data-collection period. Statistical analyses were performed using SAS 9.2 (Cary, NC). P-values were two-sided with 0.05 denoting statistical significance.

In total, 1,442 (60%) study participants reported no hypercholesterolemia, 371 (16%) reported untreated hypercholesterolemia, 514 (22%) reported treated hypercholesterolemia, and 40 (<1%) did not know their hypercholesterolemia status. Descriptive characteristics of these groups are provided in Chapter 3 (Table 3.1). Compared to those who self-reported no hypercholesterolemia, those who self-reported untreated hypercholesterolemia had an increased

prevalence of non- advanced adenomas (OR, 1.40; CI: 1.03-1.91; Table 1.1) and advanced adenomas (OR, 1.88; CI: 1.17-3.01). The association for advanced adenomas was most prominent among women (OR, 2.59; CI: 1.30-5.14) and among those who reported no previous colorectal endoscopy (OR, 2.46; CI: 1.34-4.53).

Compared to those who self-reported no hypercholesterolemia, those who self-reported treated hypercholesterolemia had an increased prevalence of advanced adenoma, but this association was not statistically significant (OR, 1.34; CI: 0.84-2.15). Among those with zenith pre-colonoscopy LDL≥160 mg/dL, however, we observed a decreased prevalence of non-advanced adenoma for those who self-reported treated hypercholesterolemia, compared to no hypercholesterolemia (OR, 0.49; CI: 0.25-0.94). These results based on participant self-report suggest that untreated hypercholesterolemia may be a risk factor for advanced adenomas, and cholesterol-controlling treatment may reduce the prevalence of adenomas among those with hypercholesterolemia.

Pharmacy-confirmed statin use and polyp prevalence

Group Health maintains a database of laboratory test results dating back to the 1980s. All available LDL measurements were extracted from at most 20 years prior to each participant's study colonoscopy. We determined each participant's highest pre-colonoscopy LDL measurement. LDL measurements were unavailable for about 30% of participants, as these were not routinely used to assess cardiovascular disease risk at Group Health until the later period of data collection. Measurements were intended to be fasting, but compliance with this requirement could not be verified.

Information on lipid-controlling drug prescriptions dispensed at eligible pharmacies, including generic name, dose, and fill-date, were extracted from electronic pharmacy records at Group Health (57). Medication types included both 3-hydroxy-3-methylglutaryl coenzyme A reductase inhibitors, commonly referred to as statins (atorvastatin, fluvastatin, lovastatin, pravastatin, rosuvastatin, simvastatin), as well as prescription non-statins (cholestyramine, colesevelam, colestipol, ezetimbe, fenofibrate, gemfibrozil, niacin). Those who used only prescription non-statins were excluded from analyses. To help avoid misclassification due to medication non-compliance, those with ≥2 separate statin prescription fills were considered to be users and those with 0 or 1 fill were considered to be non-users.

Those who ever used statins generally had a higher peak LDL (mean (SD) pre-colonoscopy zenith among controls not using statins: 135 (36) mg/dL; using statins: 167 (41) mg/dL). We estimated the association between statin use and colorectal polyps. Participants who did not know their hypercholesterolemia status (N=40) and those who used only non-statin lipid-controlling drugs were excluded from this analysis (N=17). Overall, there was no evidence to suggest that statin use was associated with colorectal polyps (OR, 0.98; CI: 0.73-1.32 for non-advanced adenoma cases; OR, 1.18; CI: 0.73-1.91 for advanced adenoma cases; and OR, 1.04; CI: 0.76-1.42 for non-adenomatous polyp cases; Table 1.2). There was little heterogeneity across sex and previous colorectal endoscopy status. Among those with zenith pre-colonoscopy LDL≥160 mg/dL, however, we observed a decreased prevalence of non-advanced adenoma comparing statin users to non-users (OR, 0.61; CI: 0.37-1.00). Duration of statin use was not evaluated.

Figure 1.1 displays the adjusted odds ratio for statin use for each of the 3 polyp case groups according to zenith pre-colonoscopy LDL. There was a trend indicating that statin use was associated with a lower relative prevalence of non-advanced adenomas at the highest

observed pre-colonoscopy LDL level (P=0.08 for a test of a non-zero slope of this line). This trend appeared to be consistent according to sex (Figure 1.2) and previous endoscopy status (Figure 1.3). Statin use also tended to be associated with a lower relative prevalence of advanced adenomas at the highest observed pre-colonoscopy LDL level among those with a previous endoscopy, but the difference in slopes displayed in Panel B and E of Figure 1.3 did not achieve statistical significance (P=0.18). If statins do, in fact, reduce the risk of small adenoma for only those with very high LDL and no prior colorectal endoscopy, differences in the inclusion criteria with respect to prior endoscopy and hypercholesterolemia severity between different studies may explain why the statin association with polyps has been inconsistent in the literature.

Summary

In summary, these results suggest that, among hypercholesterolemic adults (e.g., those who had LDL levels >160 mg/dL any time prior to colonoscopy), receipt of statin therapy may be associated with a lower odds of having a prevalent adenoma. In general, this finding is consistent with a higher relative odds of having a prevalent adenoma observed for those who self-report untreated hypercholesterolemia compared to those who self-report treated hypercholesterolemia. The possibility of uncontrolled confounding is an important limitation of these results, and motivates our use of genotype proxies as instrumental variables as described in Chapter 4.

Table 1.1. Association between self-report of hypercholesterolemia status and polyp prevalence, Group Health, 1998-2007

Self report of hyper-cholesterolemia [a]	N (%) Controls (N=976)	Non-advanced adenoma cases (N=657)	Advanced adenoma cases (N=166)	Non-adenomatous polyp cases (N=568)	Non-advanced adenoma cases vs. controls OR[b] (95% CI)	*P*	*P*[c]	Advanced adenoma cases vs. controls OR[b] (95% CI)	*P*	*P*[c]	Non-adenomatous polyp cases vs. controls OR[b] (95% CI)	*P*	*P*[c]
Total													
No	631 (66)	373 (57)	89 (55)	349 (63)	1 (Ref)		0.23	1 (Ref)		0.25	1 (Ref)		0.94
Untreated	130 (14)	114 (17)	36 (25)	94 (16)	1.40 (1.03, 1.91)	0.03		1.88 (1.17, 3.01)	0.009		1.13 (0.81, 1.57)	0.48	
Treated	193 (20)	165 (25)	38 (23)	118 (21)	1.12 (0.84, 1.49)	0.44		1.34 (0.84, 2.15)	0.23		1.11 (0.82, 1.50)	0.50	
Don't know	22 (--)	5 (--)	3 (--)	10 (--)									
Men													
No	221 (57)	172 (50)	47 (51)	138 (55)	1 (Ref)		0.88	1 (Ref)		0.89	1 (Ref)		0.11
Untreated	65 (17)	62 (18)	19 (21)	35 (14)	1.21 (0.78, 1.88)	0.40		1.27 (0.64, 2.52)	0.50		0.71 (0.42, 1.20)	0.20	
Treated	102 (26)	111 (32)	26 (28)	77 (31)	1.17 (0.80, 1.70)	0.43		1.20 (0.66, 2.19)	0.55		1.14 (0.75, 1.73)	0.54	
Don't know	8 (--)	3 (--)	2 (--)	2 (--)									
Women													
No	410 (72)	201 (65)	42 (59)	211 (69)	1 (Ref)		0.23	1 (Ref)		0.12	1 (Ref)		0.13
Untreated	65 (11)	52 (17)	17 (24)	56 (18)	1.48 (0.95, 2.32)	0.09		2.59 (1.30, 5.14)	0.007		1.52 (0.98, 2.35)	0.06	
Treated	91 (16)	54 (18)	12 (17)	41 (13)	1.03 (0.65, 1.64)	0.89		1.21 (0.53, 2.78)	0.65		0.97 (0.61, 1.54)	0.89	
Don't know	14 (--)	2 (--)	1 (--)	8 (--)									
No previous endoscopy													
No	304 (70)	176 (60)	57 (53)	165 (62)	1 (Ref)		0.06	1 (Ref)		0.17	1 (Ref)		0.64
Untreated	54 (12)	60 (20)	25 (23)	48 (18)	1.66 (1.05, 2.63)	0.03		2.46 (1.34, 4.53)	0.004		1.47 (0.90, 2.39)	0.12	
Treated	75 (17)	58 (20)	25 (24)	53 (20)	0.95 (0.59, 1.51)	0.81		1.48 (0.80, 2.72)	0.21		1.27 (0.79, 2.03)	0.32	

Don't know	7 (--)	4 (--)	2 (--)	3 (--)								
Previous endoscopy												
No	322 (63)	189 (54)	30 (57)	180 (63)	1 (Ref)		0.97	1 (Ref)		0.71	1 (Ref)		0.74
Untreated	74 (15)	54 (15)	11 (21)	42 (15)	1.22 (0.80, 1.87)	0.36		1.46 (0.63, 3.36)	0.37		0.88 (0.55, 1.40)	0.59	
Treated	115 (23)	107 (31)	12 (23)	64 (24)	1.21 (0.83, 1.75)	0.32		1.21 (0.55, 2.67)	0.64		0.97 (0.64, 1.45)	0.86	
Don't know	15 (--)	1 (--)	1 (--)	7 (--)									
Zenith LDL<160mg/dL													
No	602 (80)	333 (72)	82 (72)	324 (77)	1 (Ref)		0.47	1 (Ref)		0.35	1 (Ref)		0.53
Untreated	88 (12)	62 (13)	22 (19)	55 (13)	1.16 (0.79, 1.70)	0.44		1.53 (0.84, 2.76)	0.16		1.06 (0.71, 1.57)	0.78	
Treated	64 (8)	66 (14)	10 (9)	43 (10)	1.42 (0.92, 2.18)	0.12		0.98 (0.45, 2.15)	0.97		1.27 (0.80, 2.03)	0.31	
Don't know	16 (--)	4 (--)	3 (--)	7 (--)									
Zenith LDL≥160 mg/dL													
No	29 (15)	40 (21)	7 (14)	25 (18)	1 (Ref)		0.02	1 (Ref)		0.20	1 (Ref)		0.52
Untreated	42 (21)	52 (27)	14 (29)	36 (26)	0.99 (0.47, 2.10)	0.98		1.66 (0.46, 6.15)	0.43		0.73 (0.32, 1.65)	0.45	
Treated	129 (65)	99 (52)	28 (57)	75 (55)	0.49 (0.25, 0.94)	0.03		0.94 (0.29, 3.08)	0.92		0.59 (0.29, 1.20)	0.15	
Don't know	6 (--)	1 (--)	0 (--)	3 (--)									

Abbreviations: CI, confidence interval; OR, odds ratio.

[a]The "No" group answered "No" to the question "Has a doctor ever told you that you have high cholesterol?" The "Untreated" group answered "Yes" to the question "Has a doctor ever told you that you have high cholesterol?" and "No" to the question "Have you or are you currently taking medication for high cholesterol?". The "Treated" group answered "Yes" to the question "Has a doctor ever told you that you have high cholesterol?" and "Yes" to the question "Have you or are you currently taking medication for high cholesterol?".

[b]OR is adjusted for age at colonoscopy, sex, race, education, body mass index, nonsteroidal anti-inflammatory drug use, family history of colorectal cancer, estrogen-only use (in women), estrogen-plus-progestin use (in women), cigarette smoking, alcohol consumption, diabetes mellitus, fruit servings per day, vegetable servings per day, recreational and exercise physical activity, prior endoscopy (any time prior to approximately 2 years before study colonoscopy), and data-collection period.

[c] *P*-value for the ratio of the OR for "Treated" relative to "No" to the OR for "Untreated" relative to "No".

Table 1.2. Association between pharmacy-confirmed statin use and polyp prevalence, Group Health, 1998-2007

	N (%)									
					Non-advanced adenoma cases vs. controls		Advanced adenoma cases vs. controls		Non-adenomatous polyp cases vs. controls	
Statin use[a]	Controls (N=948)	Non-advanced adenoma cases (N=646)	Advanced adenoma cases (N=163)	Non-adenomatous polyp cases (N=554)	OR[b] (95% CI)	*P*	OR[b] (95% CI)	*P*	OR[b] (95% CI)	*P*
Total										
No	743 (78)	475 (74)	118 (73)	437 (79)	1 (Ref)		1 (Ref)		1 (Ref)	
Yes	205 (22)	171 (26)	44 (27)	117 (21)	0.98 (0.73, 1.32)	0.90	1.18 (0.73, 1.91)	0.49	1.04 (0.76, 1.42)	0.82
Men										
No	276 (72)	220 (65)	61 (67)	170 (69)	1 (Ref)		1 (Ref)		1 (Ref)	
Yes	108 (28)	121 (35)	30 (33)	77 (31)	1.13 (0.77, 1.66)	0.52	1.19 (0.66, 2.15)	0.56	1.22 (0.80, 1.87)	0.35
Women										
No	467 (83)	255 (84)	57 (80)	267 (87)	1 (Ref)		1 (Ref)		1 (Ref)	
Yes	97 (17)	50 (16)	14 (20)	40 (13)	0.80 (0.48, 1.33)	0.38	1.13 (0.47, 2.71)	0.78	0.81 (0.49, 1.34)	0.41
No previous endoscopy										
No	348 (81)	228 (79)	80 (75)	216 (82)	1 (Ref)		1 (Ref)		1 (Ref)	
Yes	84 (19)	62 (21)	27 (25)	48 (18)	0.76 (0.47, 1.23)	0.27	1.09 (0.59, 2.01)	0.79	0.88 (0.54, 1.45)	0.62
Previous endoscopy										
No	387 (76)	240 (69)	37 (71)	216 (76)	1 (Ref)		1 (Ref)		1 (Ref)	
Yes	119 (24)	108 (31)	15 (29)	68 (24)	1.16 (0.79, 1.74)	0.45	1.42 (0.63, 3.17)	0.40	1.14 (0.75, 1.73)	0.54
Zenith LDL<160 mg/dL					1 (Ref)		1 (Ref)		1 (Ref)	
No	660 (88)	384 (84)	97 (86)	369 (88)	1.08 (0.70, 1.67)	0.72	1.09 (0.53, 2.24)	0.81	1.11 (0.70, 1.76)	0.67
Yes	89 (12)	74 (16)	16 (14)	50 (12)						
Zenith LDL≥160 mg/dL										
No	83 (42)	91 (48)	21 (43)	68 (50)	1 (Ref)		1 (Ref)		1 (Ref)	

Yes	116 (58)	97 (52)	28 (57)	67 (50)	0.61 (0.37, 1.00)	0.05	0.67 (0.30, 1.51)	0.33	0.75 (0.44, 1.29)	0.30

Abbreviations: CI, confidence interval; OR, odds ratio.

[a] N=40 participants who responded "don't know" for their hypercholesterolemia status and N=17 participants who used only non-statin lipid-controlling drugs are excluded.

[b]OR is adjusted for age at colonoscopy, sex, race, education, body mass index, nonsteroidal anti-inflammatory drug use, family history of colorectal cancer, estrogen-only use (in women), estrogen-plus-progestin use (in women), cigarette smoking, alcohol consumption, diabetes mellitus, fruit servings per day, vegetable servings per day, recreational and exercise physical activity, prior endoscopy (any time prior approximately 2 years before study colonoscopy), and data-collection period.

Figure 1.1. Adjusted odds ratio of polyps comparing statin users to non-users according to pre-colonoscopy zenith low-density lipoprotein cholesterol. Associations for non-advanced adenoma cases vs. controls, advanced adenoma cases vs. controls, and non-adenomatous polyp cases vs. controls are shown in panels A-C, respectively.

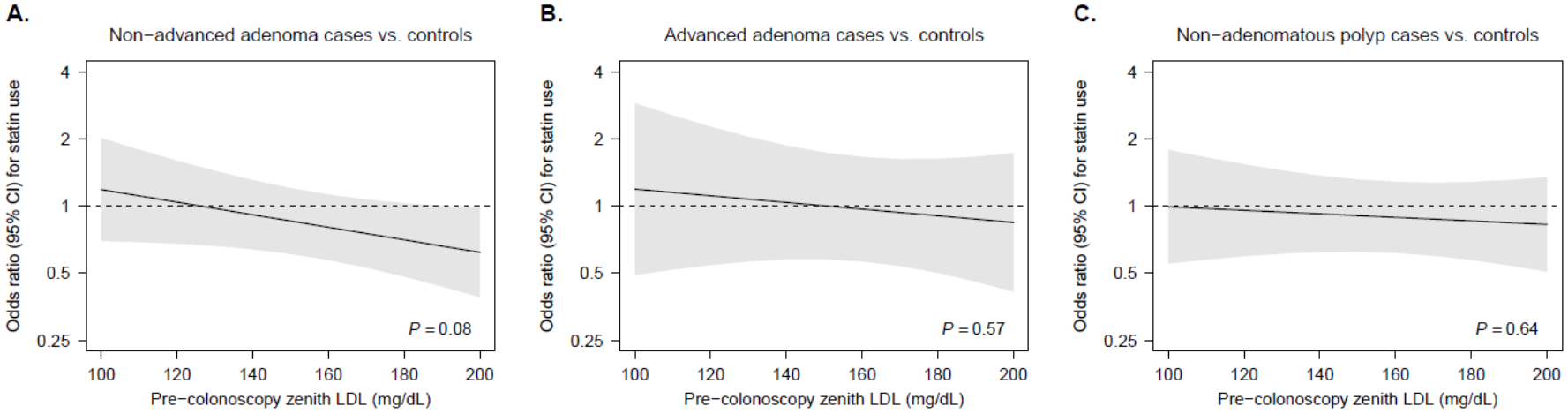


Figure 1.2. Adjusted odds ratio of polyps comparing statin users to non-users according to pre-colonoscopy zenith low-density lipoprotein cholesterol and sex. Associations for non-advanced adenoma cases vs. controls, advanced adenoma cases vs. controls, and non-adenomatous polyp cases vs. controls are shown in panels A-C, respectively, for women, and in panels D-F, respectively, for men.

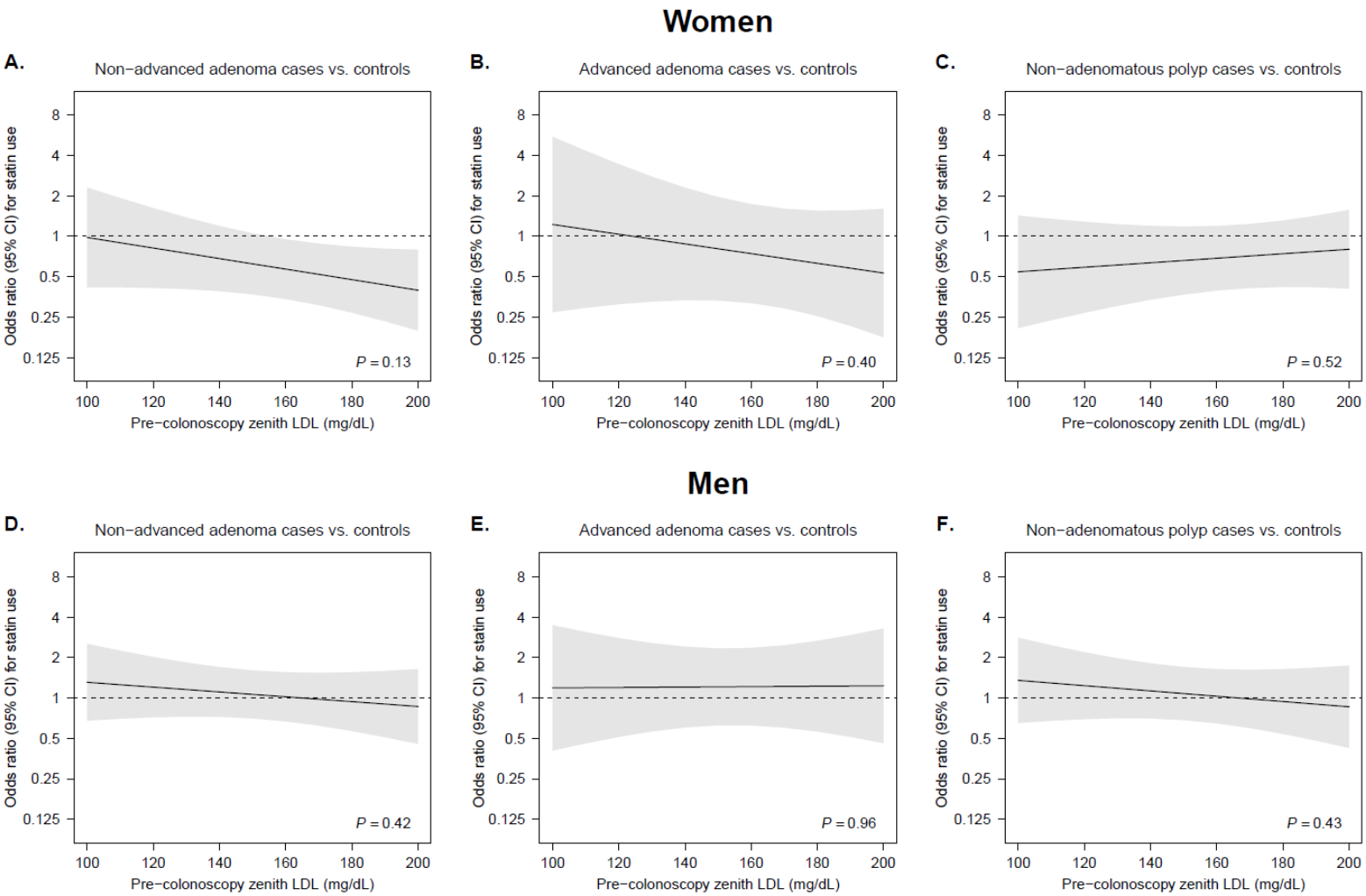
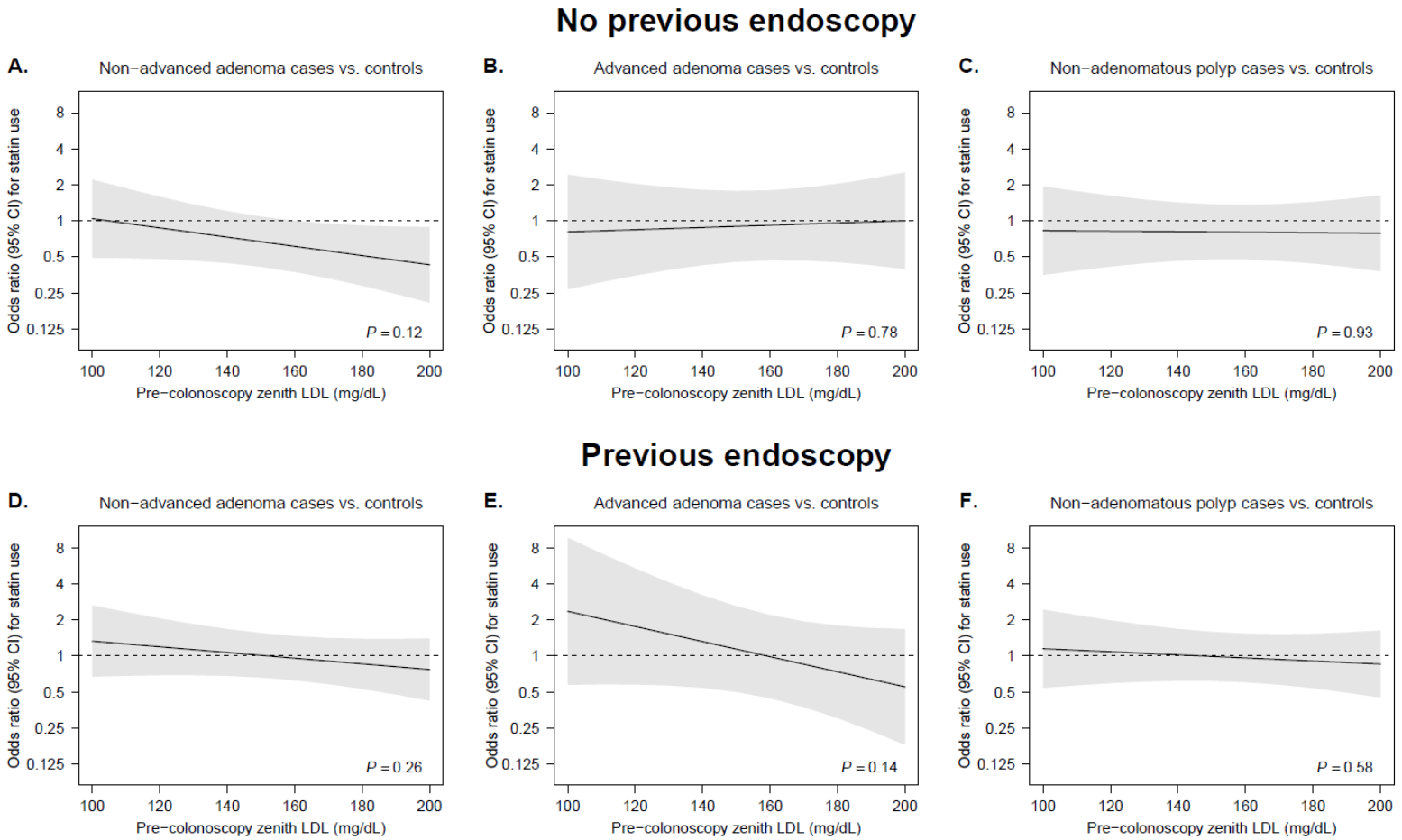


Figure 1.3. Adjusted odds ratio of polyps comparing statin users to non-users according to pre-colonoscopy zenith low-density lipoprotein cholesterol and prior endoscopy status. Associations for non-advanced adenoma cases vs. controls, advanced adenoma cases vs. controls, and non-adenomatous polyp cases vs. controls are shown in panels A-C, respectively, for those who self-reported no previous colorectal endoscopy and in panels D-F, respectively, for those who self-reported a previous colorectal endoscopy.



2. Blood lipid concentrations and colorectal adenomatous polyps: a systematic review, meta-analysis, and meta-regression of endoscopy-based case-control, 1986-2010

ABSTRACT

Objective: To inform research on cholesterol-lowering interventions and colorectal neoplasia by summarizing evidence on blood lipid concentrations at time of colorectal endoscopy among those who do and do not have adenoma.

Data sources and analyses: We systematically searched MEDLINE for colonoscopy-based or sigmoidoscopy-based case-control studies that collected blood lipid concentrations at the time of endoscopy. In eligible studies, those found to have adenoma of any size or pathology were considered cases and those found to be free from any colorectal lesion were considered controls. Included studies reported results of serum or plasma concentrations of at least one of the following lipid measurements: total cholesterol (TC), low-density lipoprotein cholesterol (LDL), high-density lipoprotein cholesterol (HDL), or triglycerides (TG).

Results: Eighteen studies conducted between 1986 and 2010 met the inclusion criteria. Combined, these studies included 6,645 adenoma cases and 21,335 normal controls. Adenoma cases had significantly higher TC than controls (adjusted mean difference (MD) = 5.13 mg/dL, 95% CI: 3.43-6.84), and higher TG than controls (MD = 18.42 mg/dL (95% confidence interval, CI: 11.71-25.12). HDL levels were 2.10 mg/dL (95% CI: 1.37-2.83) lower in individuals with adenoma compared to controls, but LDL was not significantly different (MD = 1.37, 95% CI: -1.67-4.41). Calendar year of endoscopy, percentage of men in the study, the difference in mean age between cases and controls, and whether the study excluded individuals that had undergone previous polypectomy were associated with heterogeneity in random-effects meta-regression analyses.

Conclusion: Endoscopy studies found that individuals with colorectal adenoma were more likely to have higher TC, higher TG, and lower HDL than those without adenoma.

INTRODUCTION

Colorectal adenomas are pre-neoplastic lesions that can develop into adenocarcinoma (58). It is unclear if dyslipidemia is a risk factor for adenomas, but blood cholesterol levels may be indicative of colonic exposure to bile acids (59, 60), steroid hormones (61, 62), and certain fiber metabolites (63, 64), each hypothesized to play a role in the development of colorectal neoplasia. Over the last three decades, a number of studies have investigated the association between blood lipid concentrations and risk of colorectal adenoma. Many studies reported on total cholesterol (TC), and some studies, but not all, additionally measured high-density lipoprotein cholesterol (HDL), low-density lipoprotein cholesterol (LDL), and triglycerides (TG). Recent findings on the preventive potential of statins (3-hydroxy-3-methylglutaryl-coenzyme A reductase inhibitors) (65, 66), has renewed interest in characterizing dyslipidemia as a risk factor. Studies of the association between statin use and adenoma risk, however, have been inconsistent (67-69). In order to inform research in this area, and quantify the extent to which blood cholesterol concentrations are associated with adenoma prevalence, we conducted a systematic review and meta-analysis of peer-reviewed publications from 1986-2010. All included studies were conducted in settings where use of cholesterol-lowering medications was uncommon.

METHODS

Literature search

We searched MEDLINE for original research articles published between the first available indexing date (January 1, 1966) and March 1, 2011 using the following search terms: (adenoma OR polyp OR hyperplastic) AND (cholesterol OR lipid OR triglyceride) AND (plasma OR serum OR blood) AND (colorectal OR colon OR rectal OR rectum). Our search was limited to studies involving humans, published in English, and did not include proceedings from scientific conferences. The reference lists of all full text articles were hand-searched for additional studies that may have been omitted in the database search.

Included studies were case-control studies where all participants underwent endoscopy (sigmoidoscopy or colonoscopy) for any indication (e.g., individuals found to have colorectal adenomas of any size or pathology were considered cases, and those found to be free from any colorectal lesion were considered controls). Studies were required to report a serum or plasma concentration of at least one of TC, HDL, LDL or TG from a fasting blood sample taken within 48 hours of the endoscopy visit. We did not include non-adenoma outcomes in studies that additionally reported diagnoses of hyperplastic polyps or invasive adenocarcinoma.

Studies were excluded if: 1) there were fewer than 30 adenoma cases or 30 normal controls; 2) it was unclear if all controls were screened negative of any colorectal lesion; 3) individuals with a personal history of colorectal cancer or colectomy were not excluded from participation; 4) the lipid assessments were not at the time of endoscopy or were not from blood (e.g., documented retrospectively from medical records or estimated from interview or semi-quantitative food frequency questionnaire); 4) we suspected possible reporting errors; or 5) the study sample was restricted to non-generalizable populations with specific comorbidities (e.g., alcoholics). Quantitative assessment of study quality was not performed. All search and eligibility assessments were made by the author (MNP).

Data abstraction

We abstracted the following data from eligible studies: range of calendar years of endoscopy, number of adenoma cases, number of endoscopy-negative controls, mean and standard deviations (SD) of TC, HDL, LDL, TG, difference in mean age between the case group and the control group, the proportion of men in the study, endoscopy type (colonoscopy or sigmoidoscopy), and whether the study excluded individuals that had undergone previous colorectal polypectomy. Studies that did not specifically report whether those with a history of previous polyps were ineligible were assumed to have not made such exclusions.

When abstracting data from articles, we made a limited number of assumptions to ensure consistency across studies. Means and SD were approximated from medians and ranges using the approximation suggested by Hozo et al. (70). If only sex-specific means were reported, we derived a pooled mean by taking the arithmetic average weighted by the number of men and women. Means and SD were approximated from data presented categorically by assuming uniformly distributed data within each category. If blood lipid concentration data were normalized by log-transformation, we used exponentiated means and SD. HDL and LDL concentrations reported in mmol/L were converted to mg/dL by dividing by 0.02586. TG concentrations reported in mmol/L were converted to mg/dL by dividing by 0.01129.

Statistical analyses

Differences in the mean blood concentration in mg/dL for each lipid component (mean for controls subtracted from the mean for cases) was pooled using DerSimonian-Laird random-effect models with inverse variance weighting (71). Separate analyses were performed for TC, HDL, LDL, and TG. Between-study heterogeneity was quantified using Cochran's Q statistic and I^2 (72). We assessed the possibility of publication bias by visual inspection of funnel plots

that displayed the standard error of the difference in means plotted against the value of the difference in means, and formally tested for asymmetry using Egger's regression test (73).

Study-specific adjusted relative risks reported for adenomas for each trait were plotted using circles proportional to the precision (inverse variance) of the estimate. Studies that presented effect estimates for quartiles appear as three connected circles. Studies that presented effect estimates for quintiles appear as four connected circles. Studies that presented effect estimates for a binary exposure variable (e.g., triglycerides ≥ 150 vs < 150 mg/dL) appear as a single circle. Points are plotted at the mid-point of the category. For open-ended categories, the points are plotted 25 mg/dL beyond the cut-point for LDL and triglycerides and 10 mg/dL beyond the cut-point for HDL.

Random effects meta-regression (74) was used to estimate pooled effect estimates that adjusted for four study-specific covariates: calendar year of endoscopy (mid-point of range of years; continuous), the proportion of men in the study (continuous), difference in mean age (years) between case group and control group (continuous), and whether individuals with previous polypectomy were excluded from the study (yes or no). Adjusted pooled effects were calculated as the estimated marginal means of the random-effects meta-regression function based on the average of each predictor variable weighted in the same manner as in the meta-analysis. As a sensitivity analysis, we evaluated all pooled estimated from a leave-one-out analysis, and also report pooled effect estimates for fixed-effects models. All meta-analyses and meta-regression was performed using the "metafor" package for R 2.13.0 (R Foundation for Statistical Computing; Vienna, Austria). All statistical tests were two-sided, with $P \leq 0.05$ considered statistically significant.

RESULTS

Description of included studies

Our database search resulted in 121 potential articles published between October 1974 and January 2011. Each of the 121 study abstracts were assessed for eligibility, and 91 were determined to be clearly ineligible based on study design inclusion criteria. After complete assessment of the full text of 30 potentially eligible studies, 12 were excluded (Figure 2.1). The 18 studies included in the quantitative data synthesis involved a total of 6,645 adenoma cases and 21,335 controls and were conducted between 1986 and 2010. Five studies measured each of the four blood lipid components of interest, and four studies measured three of the four components. Overall, 15 of 18 studies measured TC and TG, 12 of 18 measured HDL, and 6 of 18 measured LDL.

Half of all included studies were published since 2000, and all of the studies from 2000-2010 were in Asian populations (Table 2.1). The age range of included participants differed for each study, but was typically between 40 and 70 years. The majority of the included studies sampled an unmatched series of endoscopy patients; five studies attempted to reduce the impact of potential confounding factors by matching cases to controls on at least one characteristic (age). As expected, adenoma cases were substantially older than normal controls in studies that did not match on age (among the unmatched studies, the mean age of adenoma cases exceeded the mean age of controls by an average of 6.2 years). One study (Shinomiya et al. (39)) did not specifically report mean age, but because there was little variation in age (all participants were between 47 and 55), we assumed no difference in mean age between cases and controls for this particular study.

Six studies included only men. In the 12 studies with both sexes, slightly more men than women were enrolled. Two studies did not specifically report on whether previous polypectomy was considered, and were assumed not to have made such an exclusion. Consistent with the

prevalence of adenoma among average risk individuals (75), controls typically outnumbered adenoma cases 5-to-1 in unmatched studies. It was not always clear how controls were sampled. For example, in Tabuchi et al. over 80% of all participants were adenoma cases (76). None of the studies reported on the prevalence of use of cholesterol-lowering medications (e.g., statins), which were not as commonly used in America and Europe prior to the late 1990s and are not commonly prescribed in Asian populations.

Blood lipid concentrations

Mean TC, HDL, LDL, and TG ranged from 189.2-237.9, 44.8-59.6, 108.7-140.6, 101.0-174.3 mg/dL in normal controls, and 190.2-243.8, 44.4-58.2, 105.0-154.3, and 113.0-229.3 mg/dL in adenoma cases, respectively. Compared to the European and American studies, those conducted in Asian populations generally had lower TC, LDL, and TG in adenoma-free controls. One American study, (Bird et al. (9)) consistently had the highest mean TC, LDL, and TG blood concentrations in controls across all 18 studies. This study, however, also had the highest mean blood HDL level.

Meta-analyses

Figures 2.2-2.5 display the pooled difference in mean blood lipid concentration (mg/dL) between adenoma cases and normal controls for TC, HDL, LDL, and TG, respectively. Adenoma cases had significantly higher TC and TG levels, and lower HDL levels compared to controls. Most prominently, the TG concentration of individuals with adenoma was 18.42 mg/dL (95% CI: 11.71-25.12) higher than that of controls ($P < 0.0001$). TC in adenoma cases was 5.13 mg/dL (95% CI: 3.43-6.84) higher than that of controls ($P < 0.0001$). In contrast, HDL levels were 2.10 mg/dL (95% CI: 1.37-2.83) lower in individuals with adenoma compared to controls ($P <$

0.0001). The pooled mean difference in LDL between the two groups was not statistically significant.

We observed substantial heterogeneity between studies. For all four blood lipid measures, the null hypothesis of no heterogeneity was rejected based on Cochran's *Q* at the 0.01 level of significance (I^2 ranged from 58.5% (95% CI: 21.4-78.1%) for HDL to 76.8% (95% CI: 61.2-86.1%) for TG). In leave-one-out analyses, the estimated pooled mean difference (95% CI) from random-effects models were not materially different (from 4.17 (1.59, 6.76) to 5.68 (2.87, 8.49) for TC, from -1.84 (-2.74, -0.94) to -2.39 (-3.27, -1.52) for HDL, ranged from -1.41 (-4.69, 1.88) to 2.29 (-3.71, 8.29) for LDL, and from 15.97 (10.58, 21.35) to 19.75 (13.18, 26.32) for TG). Pooled mean differences from fixed-effects models did not substantially differ from the estimated combined effect of the random-effects analyses: the estimated pooled mean difference (95% CI) from fixed-effects models were 4.95 (3.65, 6.24), -2.25 (-2.77, -1.73), 1.11 (-1.89, 4.11), and 17.37 (14.75, 19.99) for TC, HDL, LDL, and TG, respectively. We did not find evidence of publication bias for any of the four lipid measures (Figure 2.6), as no statistically significant departures from funnel plot symmetry were detected from Egger's test. Adjusted relative risks for adenomas are displayed in Figure 2.7. Adjustment variables differ by study and are included in Table 2.1. The solid square represents the reference group, and the size of the circles is proportional to the precision (inverse variance) of the estimate.

Meta-regression

There were enough degrees-of-freedom for each lipid measurement to perform meta-regression with three continuous covariates (year, sex, age) and one binary covariate (previous polypectomy) (Table 2.2). More recent case-control studies tended to have a smaller mean difference in TC (P = 0.03). The mean difference in TC was also significantly decreased with

increasing proportion of men in the study ($P = 0.01$). The mean difference in HDL decreased as the difference in mean age between cases and controls increased ($P = 0.01$). Studies that did not restrict attention to polyp-naïve individuals showed a larger mean difference in TG than those studies that excluded individuals with previous polyps ($P = 0.04$). Collectively, the four covariates accounted for a significant portion of heterogeneity for TC, HDL, and LDL, but not TG (partitioned Q for the residual heterogeneity in the full model was $Q = 8.78$ (df = 10, $P = 0.55$) for TC, $Q = 8.57$ (df = 10, $P = 0.28$) for HDL, $Q = 0.54$ (df = 1, $P = 0.46$) for LDL, and $Q = 34.5$ (df = 9, $P < 0.0001$) for TG). Consequently, adjustment resulted in increased precision for the mean difference from random-effects meta-regression models compared to the unadjusted effect sizes for TC, HDL and LDL, but not TG (Figures 2.2-2.5).

DISCUSSION

Despite the fact that lipid levels and adenoma have been studied for decades, to our knowledge, these results have not previously been synthesized in a meta-analysis. Our meta-analysis indicates that individuals with colorectal adenomas have higher levels of TG and lower levels of HDL than those without adenoma. The large magnitude of the association with TG is likely driving the significant increase in TC that was observed for cases relative to controls. In meta-regression for TG, calendar year, age, and sex were not significant predictors of the mean difference. Moreover, it is unlikely that an unmeasured confounding variable would entirely account for an effect size of this magnitude. The meta-regression results did, however, suggest that polyp-naïve study populations were less likely to exhibit as pronounced a difference in TG concentration between adenoma cases and controls than studies that included individuals with a history of adenoma. Whether this indicates a relation between metachronous ademona and elevated triglycerides is unclear.

It should be noted that our estimated pooled effect for LDL was based on only 6 of the 18 studies (1,236 cases and 2,515 controls). The limited number of studies that measured LDL was unexpected given that it is the primary therapeutic target for cardiovascular disease prevention. Statins have been clearly demonstrated to lower serum LDL concentrations, but have limited ability to increase HDL and or lower TG (77). Many studies of statins and colorectal neoplasia make reference to the study by Bayerdorffer et al. that observed individuals with adenoma had LDL levels that significantly exceeded that of lesion-free controls (10). This result, however, is not supported by other studies, all of which reported no association.

Our meta-analysis has several limitations. All of the included studies were designed as cross-sectional case-control studies with blood lipid evaluations made at the same time as the determination of adenoma diagnosis. While the common design provides some consistency upon which to pool results, the limitations of this design prevent us from establishing the temporality between exposure and outcome. There is a notable lack of longitudinal endoscopy studies with serial lipid measurements (78). We chose to pool the crude mean blood lipid concentrations in the case group and the control group, which were consistently reported in each included study. Any associations between these unadjusted mean differences and the prevalence of adenoma, though, are likely confounded by a number of risk factors including age, sex, BMI, smoking history, and other features of the metabolic syndrome.

Nearly all of the 18 studies in our meta-analysis reported adjusted odds ratios (OR) for the association between diagnosis of adenoma and contemporaneous blood lipid concentrations. Adjusted ORs from logistic regression were difficult to combine as several different categorizations of lipid levels were used (e.g., fifth quartile vs. first quartile, third tertile vs. the first tertile, above and below some cut-off in mg/dL), and each analysis chose to adjust for a different set of confounders. The units of the OR and the adjustment variables not only impact

interpretation of the measure of association, but also affect the precision of the estimate. The differences in precision have substantial implications in calculating pooled estimates that are weighted by the inverse of the variance. We attempted to account for some of the potential confounding factors using meta-regression, but few studies summarized measures of relevant variables, and the few available degrees-of-freedom limited the number of covariates that could be considered in the meta-regression models.

In summary, these results suggest that high TG and low HDL

Figure 2.1. Flow diagram illustrating the study selection process for endoscopy-based case-control studies of the association between cholesterol and adenoma prevalence.

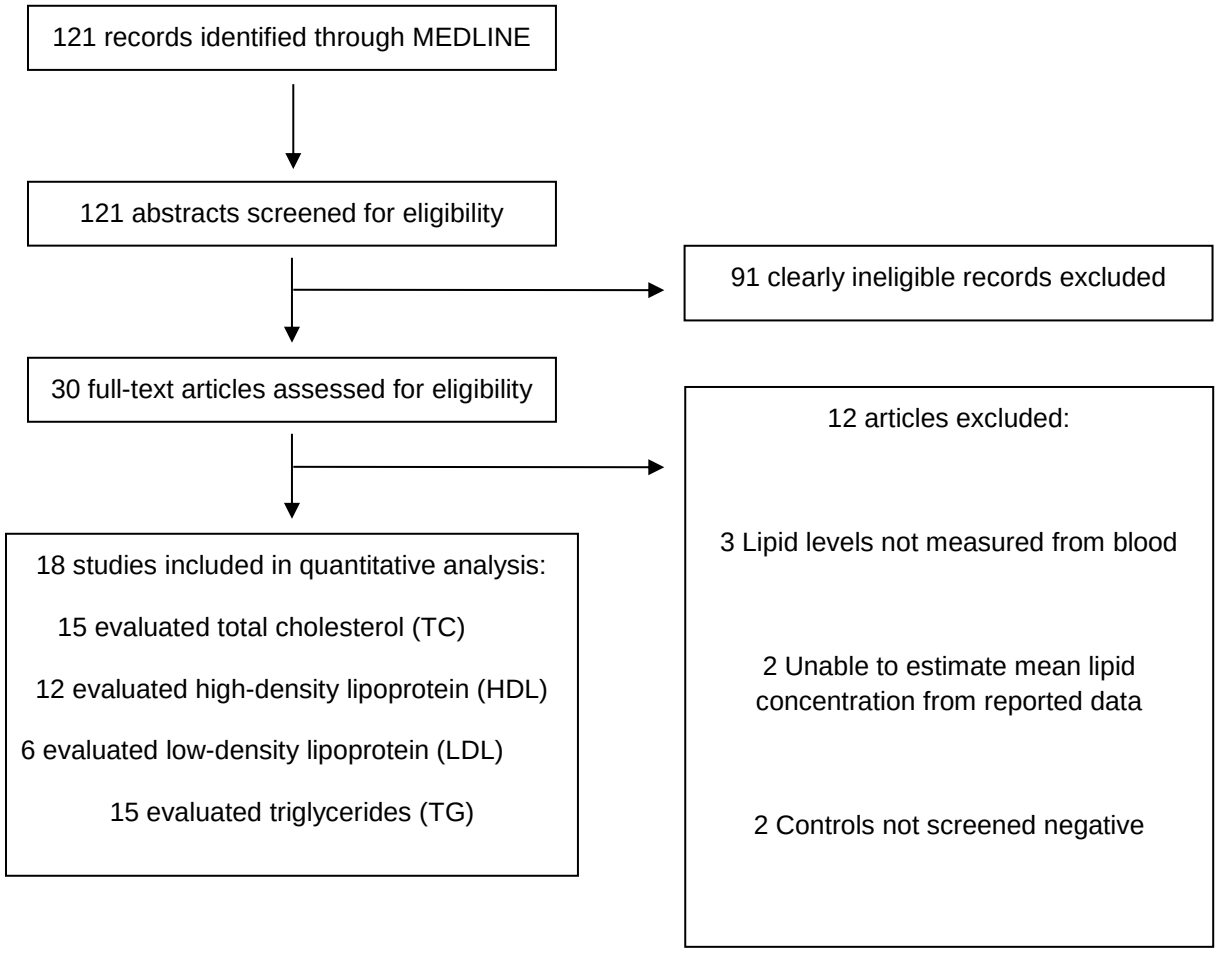


Table 2.1. Selected characteristics of 18 endoscopy-based case-control studies of the association between cholesterol and adenoma prevalence.

Study	Country	Years of Endoscopy	No. Adenoma Cases	No. Controls	Adjustment for Potential Confounding Factors	% Men	Age Range (years)	Excluded History of Polyps	Endo-scopy Type	Age Difference[a] (years)
Berry et al., 1986 (79)	Israel	1982-1985	31	60	NR	50.0	NR	N	C	5.2
Mannes et al., 1986 (80)	Germany	1982-1983	155	687	Age, BMI	46.4	34-87	N	C	10.2
Demers et al., 1988 (81)	USA	1981-1982, 1985	94	1,134	Age	100	18-70	N[d]	S	12.0
Kono et al., 1990 (6)	Japan	1986-1988	88	1,055	BMI, smoking, alcohol, TC, HDL, TG	100	49-56	Y	C	0
Bayerdorffer et al., 1993 (A) (11)	Germany	1988-1989	194	628	None	46.8	NR	N	C	10.8
Bayerdorffer et al., 1993 (B) (10)	Germany	1987-1988	288	1,192	NR	46.8	NR	N	C	11.7
Kono et al., 1993 (7)	Japan	1988-1990	138	909	BMI, smoking, alcohol	100	48-56	Y	S	0
Bird et al., 1996 (9)	USA	1991-1993	486	520	Age, sex, year, center[b]	66.0	50-75	Y	S	0.2
Manus et al., 1997 (5)	Germany	1990-1991	146	519	Age, alcohol	61.0	50-60	Y	S	0.5
Park et al., 2000 (8)	Korea	1997-1998	134	134	Age, BMI, education, family history, alcohol	100	NR	N	C	0
Shinomiya et al. 2001 (39)	Japan	1995-1996	179	219	NR	100	47-55	Y	C	0[e]
Wang et al., 2005 (82)	Taiwan	2001-2002	341	4,122	Age, sex	58.6	> 20	N	S	6.4
Chung et al., 2005 (83)	Korea	2002-2004	105	105	Age, sex, BMI, glucose, TG, TC	56.0	35-75	N[d]	C	-0.1
Tabuchi et al., 2006 (76)	Japan	1995-2003	3,920	954	Age, sex, TG, TC	61.3	10-94	N	C	15.3
Otani et al., 2006 (84)	Japan	2004-2005	782	738	Age, sex, year[c]	66.3	40-80	Y	C	0.8
Lee et al., 2008 (3)	Korea	2005	689	1,209	Age, education, income, smoking, alcohol, exercise, medications, BMI, waist circ.	100	40-70	Y	C	2.5
Kang et al., 2010 (2)	Korea	2006-2007	1,122	1,122	Age, sex, smoking, alcohol, family history, NSAIDs	77.2	40-75	Y	C	0
Liu et al., 2010 (4)	China	2006-2008	719	3,062	Age, sex, smoking, alcohol	57.4	> 30	Y	C	6.2

[a] Mean age in cases minus mean age in controls
[b] includes 89 unmatched controls and 55 unmatched cases
[c] matching only performed for women
[d] When it was not reported if those with previous polyps were excluded, it was assumed they were not
[r] Assumed to be no difference in mean age based on narrow age range of participants
Abbreviations: BMI, body mass index; C, colonoscopy; HDL, high-density lipoprotein cholesterol; N, No; NR, not reported; NSAIDS, nonsteroidal anti-inflammatory drugs; S, sigmoidoscopy; TC, total cholesterol; TG, triglycerides, Y, yes

Figure 2.2. Random-effects pooled mean difference in total cholesterol (mg/dL) between adenoma cases and controls from case-control endoscopy studies, 1986-2010. Adjusted effect is based on random-effects meta-regression model with adjustment for year, sex, age, and previous polypectomy. CI, confidence interval; SD, standard deviation.

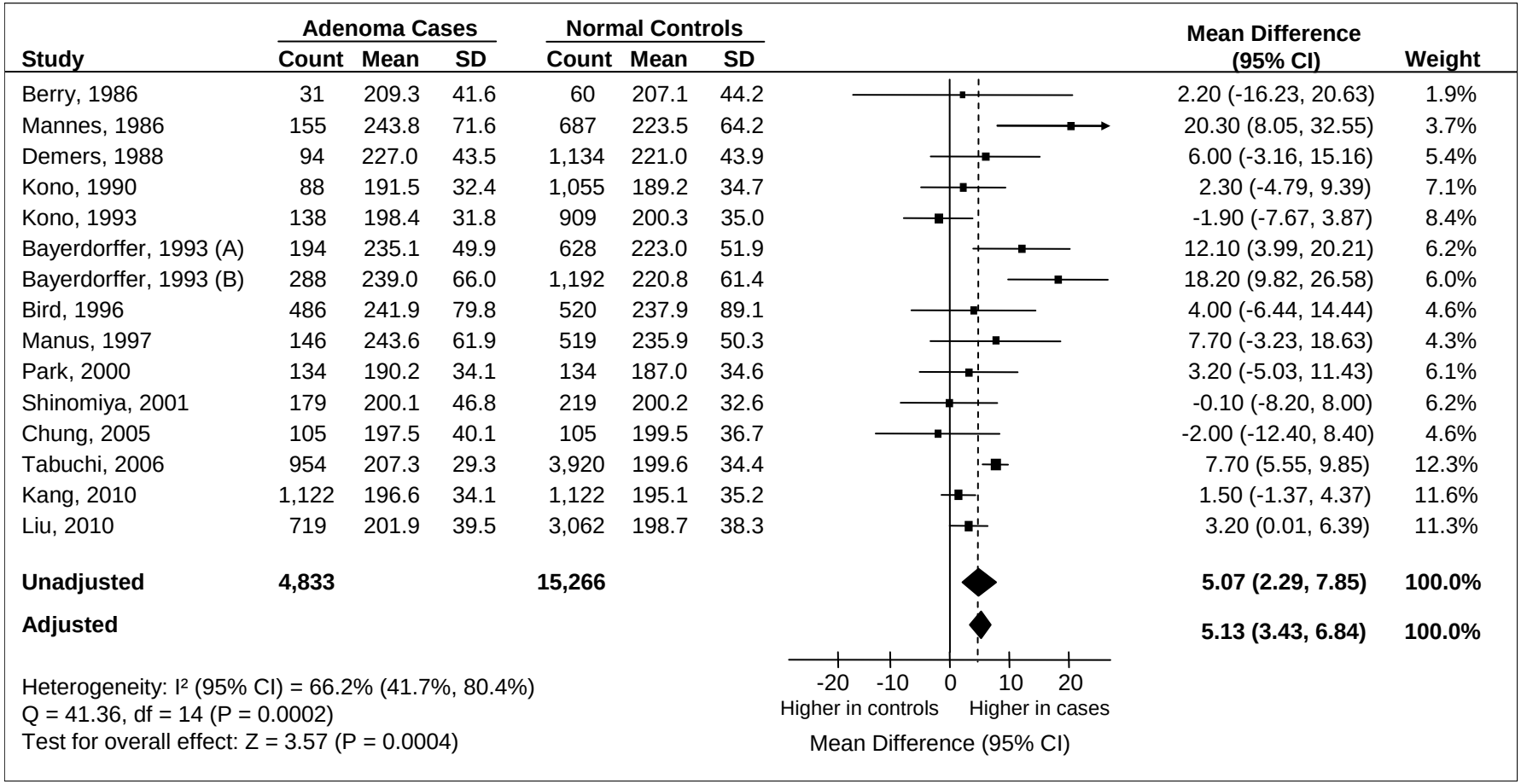


Figure 2.3. Random-effects pooled mean difference in HDL (mg/dL) between adenoma cases and controls from case-control endoscopy studies, 1986-2010. Adjusted effect is based on random-effects meta-regression model with adjustment for year, sex, age, and previous polypectomy. CI, confidence interval; SD, standard deviation.

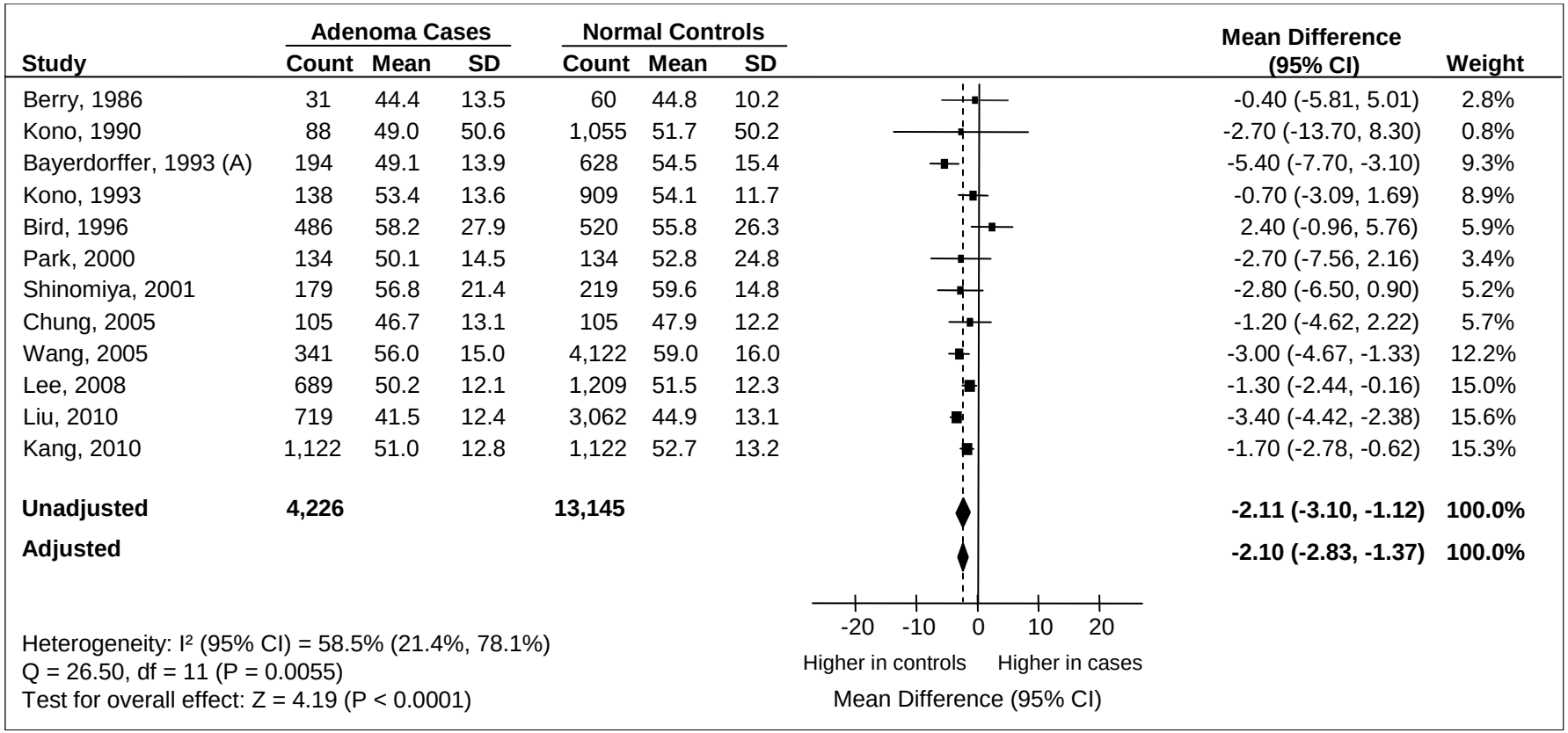


Figure 2.4. Random-effects pooled mean difference in LDL (mg/dL) between adenoma cases and controls from case-control endoscopy studies, 1986-2010. Adjusted effect is based on random-effects meta-regression model with adjustment for year, sex, age, and previous polypectomy. CI, confidence interval; SD, standard deviation.

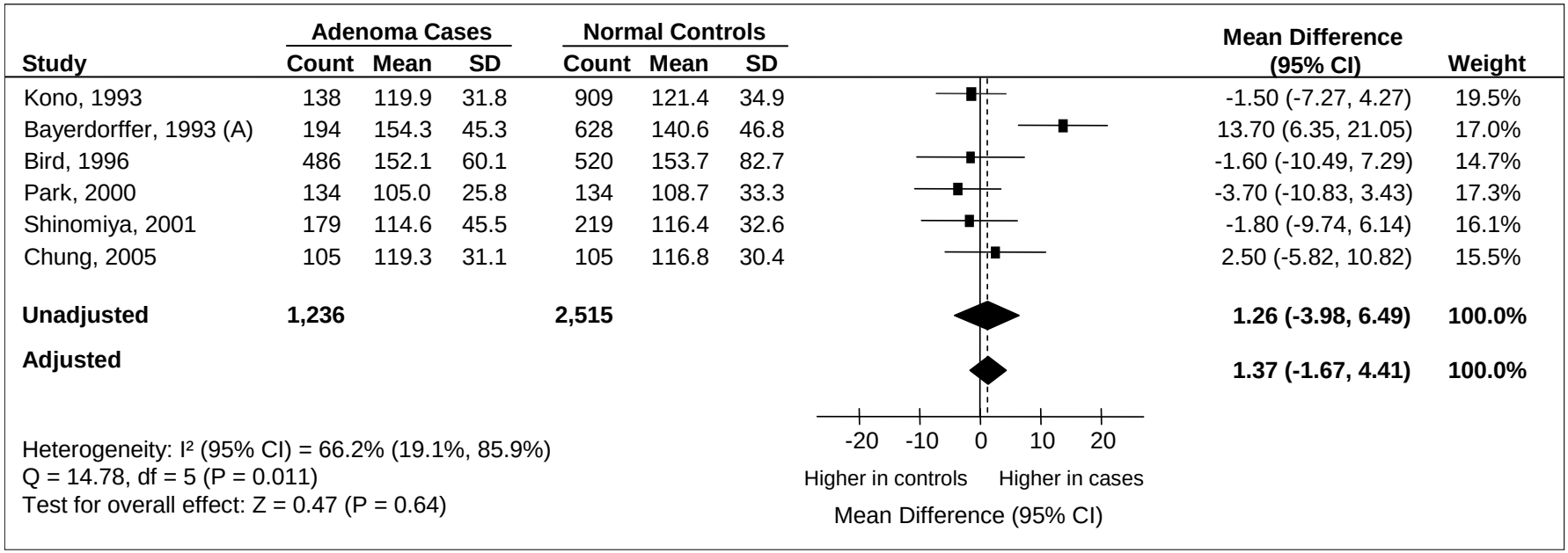


Figure 2.5. Random-effects pooled mean difference in TG (mg/dL) between adenoma cases and controls from case-control endoscopy studies, 1986-2010. Adjusted effect is based on random-effects meta-regression model with adjustment for year, sex, age, and previous polypectomy. CI, confidence interval; SD, standard deviation.

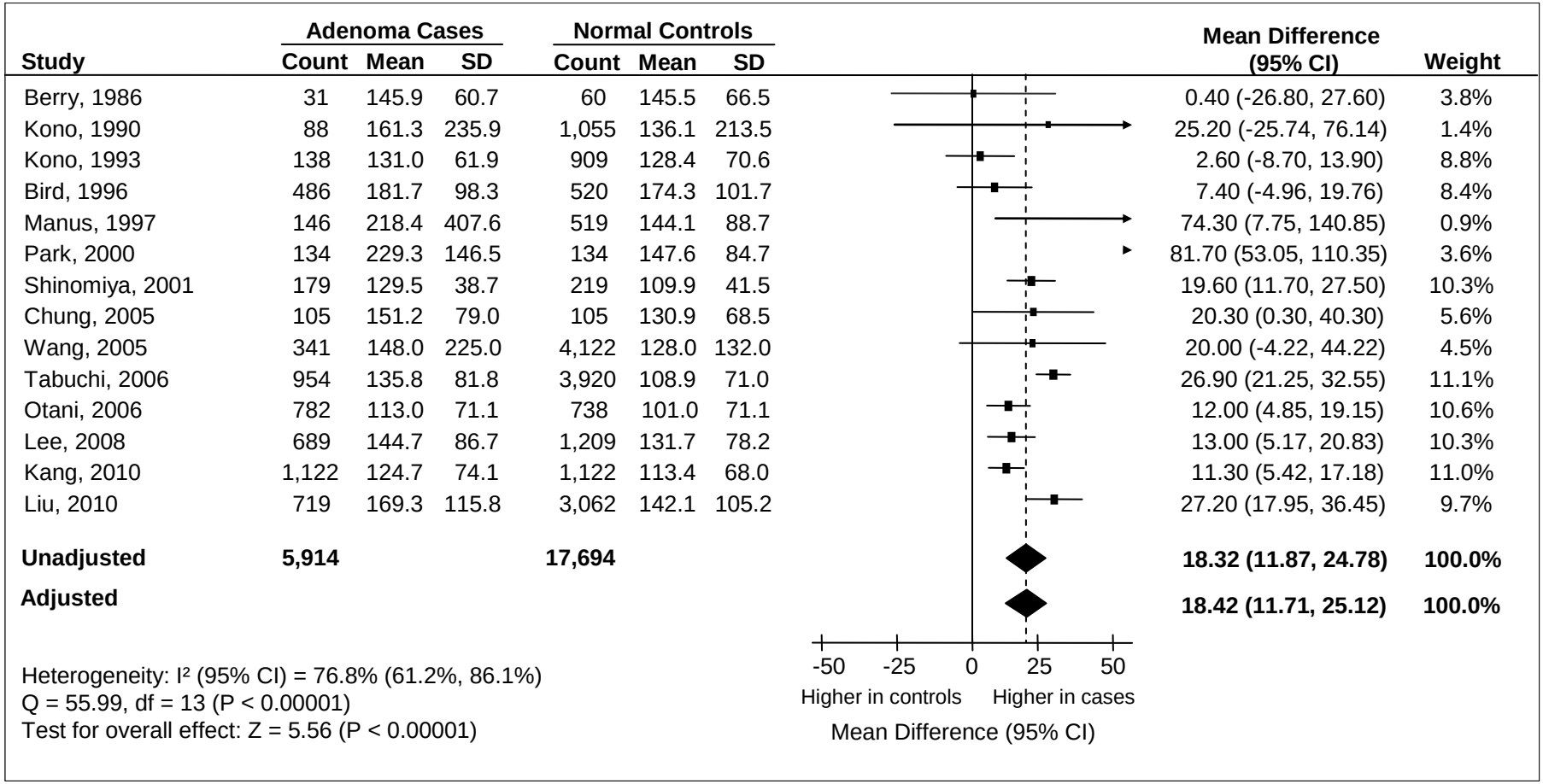
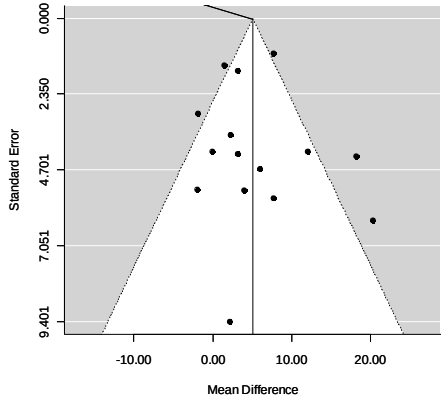


Table 2.2. Results of random-effects meta-regression analyses. Characteristics include year of endoscopy, percentage of men in the study, difference in mean age between the case and control groups, and whether the study excluded individuals that has a previous polypectomy.

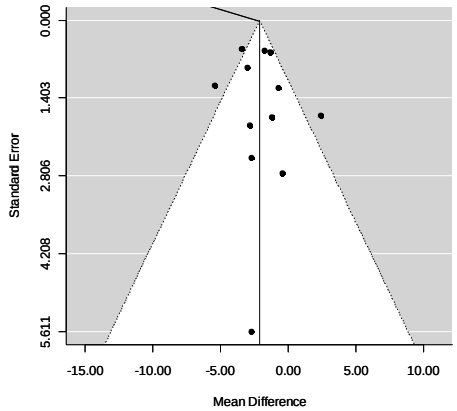
	TC (N Studies = 15)		HDL (N Studies = 12)		LDL (N Studies = 6)		TG (N Studies = 14)	
Variable	**MD (95% CI)**	***P***	**MD (95% CI)**	***P***	**MD (95% CI)**	***P***	**MD (95% CI)**	***P***
Year of endoscopy[a]	-0.26 (-0.48, -0.03)	0.03	-0.05 (-0.16, 0.07)	0.41	0.15 (-1.28, 1.57)	0.84	0.69 (-0.35, 1.72)	0.19
Percentage of men in study[b]	-0.15 (-0.27, -0.04)	0.01	-0.007 (-0.06, 0.05)	0.78	-0.07 (-0.30, 0.16)	0.54	0.28 (-0.15, 0.72)	0.20
Difference in mean age[c]	0.11 (-0.30, 0.53)	0.60	-0.37 (-0.66, -0.08)	0.01	1.29 (-0.86, 3.43)	0.24	-0.08 (-1.88, 1.72)	0.93
Excluded previous polypectomy[d]	-1.62 (-6.93, 3.68)	0.55	0.51 (-1.61, 2.64)	0.64	1.36 (-11.31, 14.02)	0.84	-20.34 (-39.78, -0.90)	0.04

Note: Separate meta-regression models for TC, HDL, LDL, and TG; each adjusted for all four variables.
[a]per one year increase
[b]per one percentage point increase
[c]per one year increase
[d]yes vs. no (reference)
Abbreviations: CI, confidence interval; HDL, high-density lipoprotein cholesterol; LDL, low-density lipoprotein cholesterol; MD, mean difference; TC, total cholesterol; TG, triglycerides

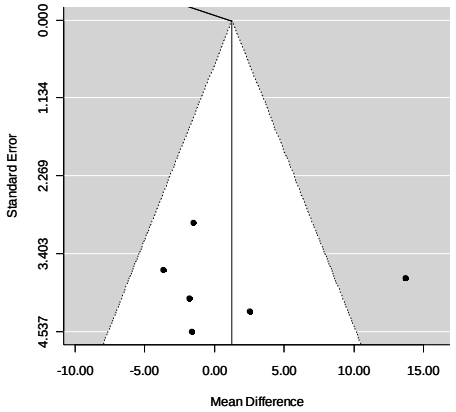
Figure 2.6. Funnel plots to assess publication bias among endoscopy-based case-control studies of the association between cholesterol and adenoma prevalence stratified by blood lipid measure. TC, total cholesterol; HDL, high-density lipoprotein cholesterol: LDL, low-density lipoprotein cholesterol; TG, triglycerides.



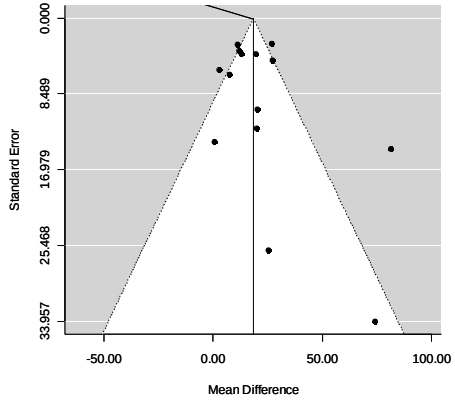
TC



HDL

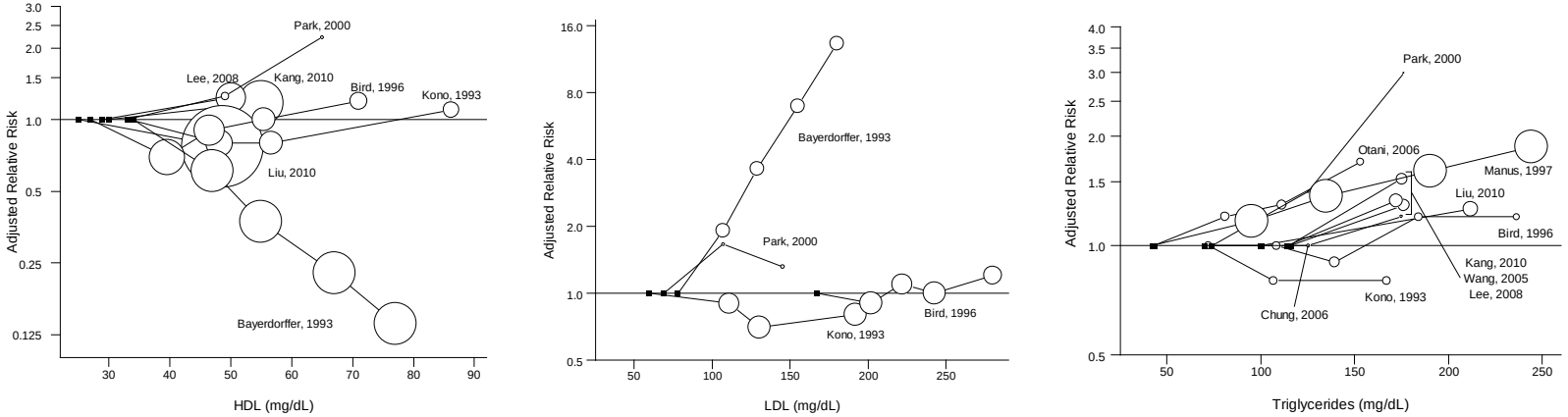


LDL



TG

Figure 2.7. Adjusted relative risks for adenomas according to HDL, LDL, and TG. Adjustment variables differ by study and are included in Table 2.1. The solid square represents the reference group. Studies that presented effect estimates for quartiles appear as three connected circles. Studies that presented effect estimates for quintiles appear as four connected circles. Studies that presented effect estimates for a binary exposure variable (e.g., triglycerides ≥ 150 vs < 150) appear as a single circle. Points are plotted at the mid-point of the category. For open-ended categories, the points are plotted 25 mg/dL beyond the cut-point for LDL and triglycerides and 10 mg/dL beyond the cut-point for HDL. The size of the circles is proportional to the precision (inverse variance) of the estimate.



3. A validition study to address ambiguity from self-report of hypercholesterolemia not treated by lipid-controlling drugs

ABSTRACT

Objective: Not all of those who self-report having been diagnosed with hypercholesterolemia receive lipid-controlling drug therapy, because lifestyle changes are often recommended before resorting to pharmacologic intervention. Investigators can choose to create simple binary classifications by grouping those who report untreated hypercholesterolemia with those who report treated hypercholesterolemia, or, alternatively, with those who report no hypercholesterolemia. We assessed the validity of both choices.

Study Design and Setting: From 1998 to 2007, a total of 2,367 adult members of Group Health, a large healthcare system in Washington State, completed a questionnaire on previous physician diagnoses of hypercholesterolemia and use of lipid-controlling drug therapy. Self-report was compared to gold standards defined from clinical cholesterol measurements extracted from medical records and prescription information from pharmacy records, going back at most 20 years prior to the questionnaire.

Results: Compared to having a low-density lipoprotein cholesterol measurement ≥130 mg/dL or pharmacy-confirmed lipid-controlling drug use, grouping untreated with treated hypercholesterolemia had 77% (95% confidence interval; CI: 74%-79%) sensitivity and 82% (CI: 79%-85%) specificity. Grouping untreated with no hypercholesterolemia had 56% (CI: 53%-60%) sensitivity and 96% (CI: 94%-97%) specificity.

Conclusion: Increased sensitivity and decreased specificity is expected when grouping those who reported untreated and treated hypercholesterolemia, but our results describe the magnitude encountered in practice. Self-report of hypercholesterolemia is fairly accurate and has higher specificity than sensitivity, regardless of how those with untreated hypercholesterolemia are classified.

Hypercholesterolemia, broadly characterized by increased blood concentrations of total cholesterol (TC), low-density lipoprotein (LDL) cholesterol, triglycerides (TG), or decreased blood concentrations of high-density lipoprotein (HDL) cholesterol, is highly prevalent (85, 86). Hypercholesterolemia is a risk factor for a number of common and fatal diseases (87), yet only about half of American adults have a desirable total cholesterol concentration (<200 mg/dL) (86).

Self-report of a personal history of hypercholesterolemia is widely used in epidemiologic research (88-96), as it can cover time periods for which measurements of blood cholesterol concentrations are unavailable. With 40 million American adults currently taking or indicated to receive statins (3-hydroxy-3-methylglutaryl coenzyme A reductase inhibitors) (97), and millions more becoming statin-eligible based on the latest clinical guidelines (97, 98), self-report of hypercholesterolemia is particularly appealing when only treatment-controlled cholesterol measurements are available. Health history questionnaires rarely ask respondents to provide quantitative estimates of previously measured cholesterol concentrations, as they often underestimated (99). Instead, a questionnaire can readily ascertain a simple binary definition of hypercholesterolemia status (yes/no).

Because self-diagnoses can be subjective (100), study participants are typically directed to report: 1) physician-diagnosed hypercholesterolemia; or 2) whether they require physician-directed treatment for hypercholesterolemia. For example, to determine hypercholesterolemia status, a questionnaire may ask: “has a doctor ever told you that you have high cholesterol?” or, alternatively, “have you ever been prescribed medication to control high cholesterol?” We use the term “treated hypercholesterolemia” throughout to refer to use of prescription lipid-controlling drugs, such as statins. We use the term “untreated hypercholesterolemia” to refer to

hypercholesterolemia that is not prescription-drug-treated, but may or may not also warrant lifestyle changes (101), such as cigarette smoking cessation, adoption of a diet low in saturated fat, and increased physical activity, or use of over-the-counter supplements, such as fish oil or plant stanols/sterols.

Exposure misclassification can lead to biased estimates of relative risk (102-104), and validation studies can be used to establish the accuracy of self-report. There have been several previous studies to validate self-report of hypercholesterolemia (99, 105-120), and, separately, others to validate self-report of lipid-controlling drug use (121-123). To our knowledge, however, validation of definitions that combine information on self-report of diagnosis and treatment is not well-described in the literature.

We ascertained self-report of prevalent hypercholesterolemia from separate diagnosis-focused and treatment-focused questions as part of a structured health history questionnaire administered during a study of Group Health members in Washington State (49). We constructed a binary variable for hypercholesterolemia status in 2 ways: 1) defining report of untreated hypercholesterolemia as "condition present" along with report of treated hypercholesterolemia (whereas report of no hypercholesterolemia represented "condition absent"); and 2) defining report of untreated hypercholesterolemia as "condition absent" along with report of no hypercholesterolemia (whereas report of treated hypercholesterolemia represented "condition present"). Both definitions were compared to gold standards derived from off-study pre-questionnaire clinical cholesterol measurements and prescription information for lipid-controlling drugs.

METHODS

Study questionnaire

Participants, 20-79 years of age, were enrollees of Group Health, a large integrated healthcare system in Washington State. All participated in a previously described colonoscopy study between 1998-2007 (49). Informed consent was obtained to access medical and pharmacy records. The study was approved by the Institutional Review Boards of Group Health and the Fred Hutchinson Cancer Research Center.

Data collection took place in 2 phases (1998-2003 and 2004-2007). A structured questionnaire was administered during an in-person interview immediately prior to the colonoscopy (1998-2003) or by telephone approximately 3 months after the colonoscopy (2004-2007). The questions were similar in both phases, and ascertained demographics and general health history. For hypercholesterolemia status, the questionnaire asked: “Has a doctor ever told you that you have high cholesterol?” (allowable answers: yes, no, don’t know). For those who answered “yes”, 2 follow-up questions were asked: 1) “When did your doctor first tell you that you had high cholesterol?” (allowable answers included an age, year, or number of years ago); and 2) “Have you or are you currently taking medication for the condition?” (allowable answers: yes, no, don’t know).

Clinical cholesterol measurements

Group Health maintains a comprehensive electronic database of laboratory test results dating back to the 1980s. We extracted all available LDL, HDL, TG, and TC measurements within 20 years prior to administration of the questionnaire. If 1 of LDL, HDL, TG, or TC was missing but the other 3 were measured on the same day, we calculated the missing value from

the Friedwald equation (124), provided that measured or calculated TG was <400 mg/dL and no calculated values was <0 mg/dL. Only a small proportion of measurements (<2%) were imputed in this manner. In the late 1990s, Group Health used TC and HDL cholesterol as the preferred method of assessing cardiovascular disease (CVD) risk (e.g., TC-to-HDL ratio). Although LDL and TG measurements could also be ordered, more TC and HDL measurements were available than LDL and TG. After adoption of National Cholesterol Education Program Adult Treatment Panel (ATP) III guidelines, which outline LDL goals in the context of other risk factors, including smoking status, hypertension, and family history (125), nearly every ordered panel included all 4 traits.

Lipid-controlling medication prescriptions

Electronic pharmacy records at Group Health include drug name and fill-date for prescriptions dispensed at outpatient pharmacies operated by Group Health (57). Participants were considered users of lipid-controlling drugs if they filled ≥2 prescriptions for any of the following prior to the study questionnaire: a statin (atorvastatin, fluvastatin, lovastatin, pravastatin, rosuvastatin, simvastatin), bile acid sequestrant (cholestyramine, colesevelam, colestipol), fibric acid (fenofibrate, gemfibrozil), cholesterol absorption inhibitor (ezetimibe), or nicotinic acid (prescription niacin; over-the-counter niacin was not assessed). Combination drugs containing statins were categorized with statins.

Definition of gold standards

Those with ≥2 separate prescription fills for lipid-controlling drugs were considered to have "true" hypercholesterolemia (those with 0 or 1 fill were considered to be free of "true" hypercholesterolemia). In addition, we defined "true" hypercholesterolemia according to 8

different cut-points are based on ATP III recommendations: 2 LDL-specific cut-points (≥130 mg/dL and ≥160 mg/dl), 2 HDL-specific cut-points (<60 mg/dL and <40 mg/dL), 2 TG-specific cut-points (≥150 mg/dL and ≥200 mg/dL), and 2 TC-specific cut-points (≥200 mg/dL and ≥240 mg/dL). For each, we required ≥2 separate measurements to meet the threshold (those with 0 or 1 measurement to meet the threshold were considered to be free of "true" hypercholesterolemia).

Exclusions

Of 2,506 participants in the colonoscopy study, we excluded 78 who did not have any available lipid measurements within 20 years of the questionnaire, and 61 who self-reported being diagnosed with hypercholesterolemia >20 years prior to the questionnaire. We excluded 75 TG measurements ≥800 mg/dL from 26 participants, as these unusually high values could result from acute toxicities (all 26 had at least 1 TG measurement <800 mg/dL on record) (126). There were 32 participants who had a lipid-controlling drug prescription prior to the first recorded blood lipid measurement (the median interval was 2.3 months; 25^{th}, 75^{th} percentile; 1.8 weeks, 10.1 months). These 32 participants were retained in primary analyses, providing a final sample size of 2,367.

Statistical analyses

Participants were classified into 4 groups based on self-report of hypercholesterolemia status: 1) no hypercholesterolemia; 2) untreated hypercholesterolemia; 3) treated hypercholesterolemia; and 4) unknown, having indicated "don't know" to having received a previous diagnosis of hypercholesterolemia or having received treatment for hypercholesterolemia. Analyses describe age at administration of questionnaire (years), sex, race (Caucasian, Black/African American, Asian/Pacific Islander, other), body mass index (BMI;

kg/m^2), cigarette smoking status (never, former, current), year of questionnaire (1998-2003, 2004-2007).

Sensitivity (proportion of truly hypercholesterolemic individuals who correctly report having the condition) and specificity (proportion of truly non-hypercholesterolemic individuals who correctly report not having the condition) were calculated with asymptotic continuity-corrected 95% confidence intervals (CI) with respect to each lipid-component-specific gold standard. Those who could not self-report their hypercholesterolemia status were excluded from the calculation of sensitivity and specificity. We assessed sensitivity and specificity for 2 binary classification choices: 1) grouping those who reported untreated hypercholesterolemia together with those who reported treated hypercholesterolemia; and 2) grouping those who reported untreated hypercholesterolemia together with those who reported no hypercholesterolemia. Let SN_1 and SP_1 denote the sensitivity and specificity, respectively, for the first choice, and SN_2 and SP_2 denote the sensitivity and specificity, respectively, for the second choice. Note that $SN_1 \geq SN_2$ and $SP_1 \leq SP_2$ (Appendix).

We also calculated attenuated odds ratio (AOR) for a hypercholesterolemia-disease association that would result from non-differential misclassification (i.e., diseased and disease-free individuals have equal sensitivity and equal specificity) from self-report using our estimates of sensitivity and specificity (127). We assumed the "true" prevalence of hypercholesterolemia among non-diseased individuals (i.e., controls in a case-control study) was equal to the estimated prevalence of hypercholesterolemia among all study participants based on our records-derived gold standards. AORs are presented for both ways of classifying untreated hypercholesterolemia, but no formal statistical test of the difference in attenuation was evaluated.

Two secondary analyses were conducted to assess the robustness of our sensitivity and specific estimates with respect to secular trends. First, we stratified by the duration of available retrospective laboratory record follow-up (duration between questionnaire and earliest available measurement of that cholesterol trait prior to questionnaire; >0 to ≤5 years, >5 to ≤10 years, >10 to ≤20 years prior to questionnaire). Second, we stratified by calendar year of administration of the questionnaire (1998-2003, 2004-2007). All analyses were performed using SAS 9.2 (SAS Institute, Inc.; Cary, NC).

RESULTS

Participant characteristics

The majority of study participants were Caucasian, and the mean age was 60 years old (93% were ≥50 years old; Table 3.1). In total, the self-reported prevalence of untreated hypercholesterolemia was 16% (N=371) and the self-reported prevalence of treated hypercholesterolemia was 22% (N=514). Those who reported treated hypercholesterolemia were more likely to be older, men, and interviewed during the later study period (2004-2007). Only 2% (N=40) of participants did not know their hypercholesterolemia status (N=35 said "don't know" regarding a diagnosis of hypercholesterolemia and N=5 reported a diagnosis but said "don't know" about lipid-controlling drug use).

Not counting those who did not know their hypercholesterolemia status, N=3 had no available HDL measurements, N=686 had no available LDL measurements, and N=627 had no available TG measurements. All participants had at least 1 TC measurement on record. Those who reported treated hypercholesterolemia had an average of 7 LDL, 12 HDL, 9 TG, and 12 TC

measurements per person, whereas those who reported untreated hypercholesterolemia had 3 LDL, 6 HDL, 3 TG, and 6 TC measurements per person, and those who reported no hypercholesterolemia had 2 LDL, 5 HDL, 3 TG, and 5 TC measurements per person. In total, 24% (N=564) filled ≥2 prescriptions for a lipid-controlling drug (84% of those who reported treated hypercholesterolemia, 7% of those who reported untreated hypercholesterolemia, and 7% of those who reported no hypercholesterolemia).

Sensitivity and specificity of self-report of hypercholesterolemia

The estimated prevalence of hypercholesterolemia based on the gold standards ranged from 34-70% depending on the trait and cut-point (Table 3.2). Classifying both self-report of untreated and treated hypercholesterolemia as "condition present" resulted in sensitivity ranging from 48-83% and specificity ranging from 73-94%. Instead, classifying only self-report of treated hypercholesterolemia as "condition present" resulted in sensitivity ranging from 30-69% and specificity ranging from 95-99%. Of the 4 traits, LDL generally had the highest sensitivity and HDL the lowest.

Consequences of non-differential misclassification from self-report of hypercholesterolemia

Based on the sensitivities and specificities we observed, both approaches for classifying untreated hypercholesterolemia resulted in rather substantial attenuation of an odds ratio for a hypercholesterolemia-disease association (Table 3.3). Differences in the degree of attention between the 2 approaches were generally small. At the higher cut-point for each trait, classifying untreated hypercholesterolemia as "condition absent" resulted in less attenuation than when classifying untreated hypercholesterolemia as "condition present". At the lower cut-point for LDL (130 mg/dL) and TC (200 mg/dL), however, there were situations when classifying

untreated hypercholesterolemia as "condition absent" resulted in more attenuation than when classifying untreated hypercholesterolemia as "condition present".

Secondary analyses

Study participants with longer duration (>10 years) of cholesterol record history at Group Health were generally older, less likely to be smokers, more likely to be lipid-controlling drug users, and had more available lipid measurements than those with shorter duration of history. Consistent with these characteristics, those with a longer duration of cholesterol record history were more likely to have hypercholesterolemia as defined by the gold standards (Supplemental Table 3.1). In general, for gold standards based on LDL, self-report had both higher sensitivity and specificity for those with more medical record history, but conclusions regarding the 2 ways of classifying those who self-report untreated hypercholesterolemia were consistent by strata.

Study participants who enrolled from 2004-2007 were also less likely to be smokers, more likely to be lipid-controlling drug users compared, and had more available lipid measurements than those who enrolled from 1998-2003. The age distribution of participants was similar in both phases of data collection. Because the TC-to-HDL ratio was used more often than a complete panel when our study began, long-term information (>10 years) on HDL and TC was more likely for those who enrolled in the later phase of data collection. Self-report always had relatively poor sensitivity for gold standards based on HDL and TC, modest improvement in sensitivity was noted in the later phase of data collection (Supplemental Table 3.2). Differences in sensitivity and specificity by calendar time, however, were not substantial enough to alter our conclusions.

DISCUSSION

In our study of adults in a managed-care setting, nearly as many participants self-reported having received a diagnosis of hypercholesterolemia not requiring lipid-controlling drugs as self-reported a diagnosis of hypercholesterolemia requiring lipid-controlling drugs. Studies may decide to classify persons who report untreated hypercholesterolemia as either "condition present" or "condition absent," thus limiting comparability of findings for studies that make different choices.

Our estimates of sensitivity and specificity were consistent with previously published values from studies that compared self-report to blood values collected as part of a prospective study (99, 105-116), and those that compared self-report to measurements documented in medical records near the time of questionnaire (117-120). On one hand, classifying untreated hypercholesterolemia with treated hypercholesterolemia may provide the best balance of good sensitivity with good specificity. On the other hand, classifying untreated hypercholesterolemia with no hypercholesterolemia tended to achieve near 100% specificity, which may be attractive for evaluations that aim to avoid false positives. For the prevalence of hypercholesterolemia we observed, we found minimal differences in the attenuation of a hypercholesterolemia-disease OR from non-differential misclassification.

The sensitivity that occurs when untreated hypercholesterolemia is considered "condition present" (SN_1) is decreased relative to the sensitivity that occurs when untreated hypercholesterolemia is considered "condition absent" (SN_2) by $\frac{n_{11}}{n_{11}+n_{21}}$, which can be thought of as the sensitivity of report of treatment for hypercholesterolemia relative to the same gold

standard of hypercholesterolemia status, among those who report a diagnosis of hypercholesterolemia. Likewise, the specificity that occurs when untreated hypercholesterolemia is considered "condition present" (SP_1) is increased relative to the specificity that occurs when untreated hypercholesterolemia is considered "condition absent" (SP_2) as a function of $\frac{n_{22}}{n_{12}+n_{22}}$, which can be thought of as the specificity of report of treatment for hypercholesterolemia, relative to the same gold standard of hypercholesterolemia status, among those who report a diagnosis of hypercholesterolemia.

The completeness and high quality of data from Group Health is a primary strength of our study (57). We assumed that information from Group Health records could serve as a gold standard for hypercholesterolemia status and adequately covered the period during which hypercholesterolemia diagnoses would occur. The prevalence of hypercholesterolemia based on our LDL gold standards was generally consistent with an age-specific prevalence reported from National Health and Nutrition Examination Survey (128).

A number of previous studies conducted among enrollees of Group Health have use cholesterol measurements from laboratory records and information on lipid-controlling drugs from pharmacy records (123, 129-132). Although some prescriptions filled outside of the Group Health system may have been missed, Boudreau et al. found that most enrollees fill prescriptions at a Group Health pharmacy, and furthermore, that the accuracy of self-report of statin use did not significantly improve when pharmacy records at Group Health were augmented with prescription data from major retail pharmacies that operate in Washington State (123).

We acknowledge several limitations. The ATP III recommendations outline LDL levels to initiate therapeutic lifestyle changes and LDL levels to initiate drug therapy within 3

categories defined by the presence of comorbidities and 10-year CVD risk predicted by the Framingham score (125). Complete ascertainment of CVD risk factors was unavailable, and we therefore did not stratify our analyses by these 3 risk categories. During the ATP III era, complete lipid panels were routine for CVD prevention and intended to be fasting. Although compliance with this requirement cannot be verified, the necessity of considering only fasting cholesterol has been recently debated (133). Evidence-based clinical guidelines for the diagnosis and management of hypercholesterolemia are constantly evolving (134, 135). With increasing physician compliance (136) and increasing adoption of statin therapy (137), the prevalence of truly untreated hypercholesterolemia has likely decreased over time. To describe changes in the accuracy of self-report over time was not a primary goal of this analysis, but we attempted to assess the influence of secular trends in secondary analyses.

Our structured questionnaire asked about several common medical conditions, including hypercholesterolemia, but a more detailed line of questioning may have limited misclassification. It is possible that ambiguity in the term "high cholesterol," particularly with respect to low HDL, may be a source of confusion. The wording of cholesterol-related questions from our questionnaire, though, was nearly identical to that of the questionnaire used by the National Health and Nutrition Examination Survey (138). Our questionnaire did not instruct participants to focus only on prescription drugs. Some participants who reported treated hypercholesterolemia may have been referring to non-prescription supplements. Consequently, our estimated prevalence of treated hypercholesterolemia from self-report may be higher than would be expected if only statin-treatment was of interest.

In general, the appropriateness of combining information from diagnosis-focused and treatment-focused questions should be carefully considered. Progressive conditions identified

from the former may be of a different severity than conditions identified from the latter (139). The sensitivity and specificity relations described here can be generalized using methods for combining sequential diagnostic tests (self report of a diagnosis can be considered a first test and receipt of treatment can be considered a second test) (140-143). Self-report of treated hypercholesterolemia (requiring a positive result on both tests) behaves like the well-characterized "believe the negative" composite test (141). Methods to formulate association-attenuation from non-differential misclassification in terms of sensitivity and specificity relations for sequential diagnostic tests may prove useful in broader applications, particularly in conjunction with approaches for dealing with imperfect standards (144). Moreover, this approach extends to any multicategory variable when there is uncertainty whether some categories should be collapsed with adjacent ones (145).

In summary, we described how combining information from diagnosis-focused and treatment-focused self-report of hypercholesterolemia can reflect evidence from clinical cholesterol measurements. Our findings may prove useful to investigators who rely on self-report to assess hypercholesterolemia status, or possibly other commonly pharmacologically-treated conditions.

APPENDIX

Refer to the cell and margin counts as labelled in Figure 1. Let SN_1 and SP_1 denote the sensitivity and specificity, respectively, when a binary variable in defined such that those who self-report of untreated hypercholesterolemia are grouped with those who self-report of treated hypercholesterolemia. Let SN_2 and SP_2 denote the sensitivity and specificity, respectively, when a binary variable in defined such that those who self-report of untreated hypercholesterolemia are grouped with those who self-report no hypercholesterolemia.

Additively: $\mathrm{SN}_2 = \frac{\mathrm{TP}_2}{\mathrm{TP}_2+\mathrm{FN}_2} = \frac{n_{11}}{n_{+1}} = \frac{\mathrm{TP}_1-n_{21}}{\mathrm{n}_{+1}} = \frac{\mathrm{TP}_1}{\mathrm{TP}_1+\mathrm{FN}_1} - \frac{n_{21}}{n_{+1}} = \mathrm{SN}_1 - \frac{n_{21}}{n_{+1}}$, and

$\mathrm{SP}_2 = \frac{\mathrm{TN}_2}{\mathrm{TN}_2+\mathrm{FP}_2} = \frac{n_{22}+n_{32}}{n_{+2}} = \frac{\mathrm{TN}_1+n_{22}}{n_{+2}} = \frac{\mathrm{TN}_1}{\mathrm{TN}_1+\mathrm{FP}_1} + \frac{n_{22}}{n_{+2}} = \mathrm{SP}_1 + \frac{n_{22}}{n_{+2}}$.

Multiplicatively: $\mathrm{SN}_2 = \frac{\mathrm{TP}_2}{\mathrm{TP}_2+\mathrm{FN}_2} = \frac{n_{11}}{n_{+1}} = \frac{n_{11}+n_{21}}{\mathrm{TP}_1+\mathrm{FN}_1} \times \frac{n_{11}}{n_{11}+n_{21}} = \frac{\mathrm{TP}_1}{\mathrm{TP}_1+\mathrm{FN}_1} \times \frac{n_{11}}{n_{11}+n_{21}} = \mathrm{SN}_1 \times \frac{n_{11}}{n_{11}+n_{21}}$,

and $\mathrm{SP}_2 = \frac{\mathrm{TN}_2}{\mathrm{TN}_2+\mathrm{FP}_2} = \frac{n_{22}+n_{32}}{n_{+2}} = \frac{n_{+2}-n_{12}}{n_{+2}} = 1 - \frac{n_{12}}{n_{+2}} = 1 - \frac{n_{11}+n_{22}}{n_{+2}} \times \frac{n_{12}}{n_{11}+n_{22}} = 1 - (1 - \mathrm{SP}_1) \times (1 - \frac{n_{22}}{n_{12}+n_{22}})$.

Let AOR_1 denote the attenuated odds ratio for a hypercholesterolemia-disease association based on SN_1 and SP_1, and let AOR_2 denote the attenuated odds ratio based on SN_2 and SP_2. For a case-control study, the odds ratio for a hypercholesterolemia-disease association based on a gold standard is $\mathrm{OR} = \frac{\pi^D(1-\pi^{\bar{D}})}{\pi^{\bar{D}}(1-\pi^D)}$, where π^D is the proportion of cases with hypercholesterolemia

and $\pi^{\bar{D}}$ is the proportion of controls with hypercholesterolemia. Note that $\pi^{D} = \frac{\pi^{\bar{D}}\mathrm{OR}}{\pi^{\bar{D}}(\mathrm{OR}-1)+1}$. As previously described (127, 146), assuming non-differential misclassification, we can express AOR_i, for $i = 1, 2$, as a function of OR and $\pi^{\bar{D}}$. That is, $\mathrm{AOR}_i = \frac{p_i^D(1-p_i^{\bar{D}})}{p_i^{\bar{D}}(1-p_i^D)}$ for $i = 1, 2$, where $p_i^{\bar{D}} = (\mathrm{SN}_i)\left(\pi^{\bar{D}}\right) + (1 - \mathrm{SP}_i)\left(1 - \pi^{\bar{D}}\right)$, and $p_i^D = (\mathrm{SN}_i)\left(\frac{\pi^{\bar{D}}\mathrm{OR}}{\pi^{\bar{D}}(\mathrm{OR}-1)+1}\right) + (1 - \mathrm{SP}_i)\left(1 - \frac{\pi^{\bar{D}}\mathrm{OR}}{\pi^{\bar{D}}(\mathrm{OR}-1)+1}\right)$. Note that $\mathrm{AOR}_i = \mathrm{OR}$ when $\mathrm{SN}_i = \mathrm{SP}_i = 1$.

Figure 3.1. Summary of binary classifications of self-report of hypercholesterolemia. Two situations are considered: 1) grouping those who reported untreated hypercholesterolemia together with those who reported treated hypercholesterolemia; and 2) grouping those who reported untreated hypercholesterolemia together with those who reported no hypercholesterolemia. All classifications are relative to a common binary gold standard.

Abbreviations: FN, false negative; FP, false positive; TN, true negative; TP, true positive.

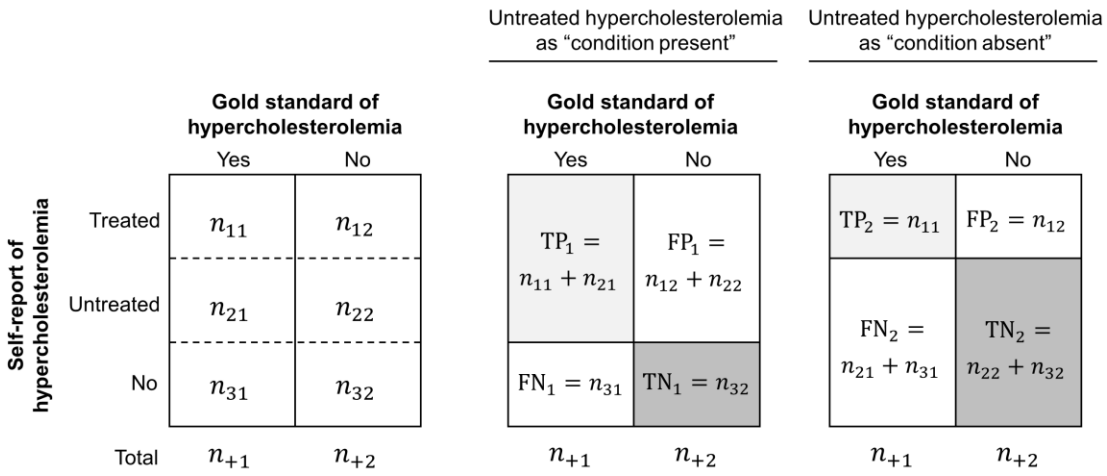


Table 3.1. Characteristics of participants according to self-report of hypercholesterolemia from study questionnaire, Group Health, 1998-2007

	Self-report of hypercholesterolemia [a]																			
	No (N=1,442)				Untreated (N=371)				Treated (N=514)				Don't know[b] (N=40)				Total (N=2,367)			
Characteristic at questionnaire administration	N	%	Mean	95% CI	N	%	Mean	95% CI	N	%	Mean	95% CI	N	%	Mean	95% CI	N	%	Mean	95% CI
Age, years			59.5	59.1, 59.9			59.6	58.8, 60.4			62.7	62.1, 63.3			64.8	61.5, 68.2			60.3	60.0, 60.6
Sex																				
Male	578	40			181	49			316	61			15	38			1,090	46		
Female	864	60			190	51			198	39			25	63			1,277	54		
Race																				
Caucasian	1,234	86			322	87			427	83			37	93			2,020	85		
Black/African American	46	3			7	2			18	4			1	3			73	3		
Asian/Pacific Islander	68	5			17	5			25	5			0	0			110	5		
Other	94	7			25	7			44	9			4	5			165	7		
BMI, kg/m^2			26.5	26.3, 26.8			27.9	27.4, 28.4			28.4	28.0, 28.8			28.0	26.3, 29.8			27.2	27.0, 27.4
Cigarette smoking status																				
Never	779	54			201	54			224	44			14	35			1,218	52		
Former	574	40			137	37			241	47			22	55			974	41		
Current	88	6			32	9			47	9			4	10			171	7		
Year																				
1998-2003	297	21			61	16			53	10			11	28			440	19		
2004-2007	1,145	79			310	84			461	90			29	73			1,927	81		

Abbreviations: BMI, body mass index; CI, confidence interval.

[a] Counts may not sum to column total because missing values are not shown.

[b] N=35 said "don't know" about a previous diagnosis of hypercholesterolemia and N=5 reported a diagnosis but said "don't know" about lipid-controlling drug use.

Table 3.2. Sensitivity and specificity of self-report of hypercholesterolemia from study questionnaire compared to gold standards based on clinical measurements and pharmacy records, Group Health, 1998-2007

Gold standard (mg/dL)[a]	Classification based on gold standard[b]				Classification of those who self-report untreated hypercholesterolemia							
					"Condition present"				"Condition absent"			
	"Condition present"		"Condition absent"		SN_1		SP_1		SN_2		SP_2	
	N	%	N	%	%	95% CI	%	95% CI	%	95% CI	%	95% CI
LDL≥130	853	52	788	48	77	74, 79	82	79, 85	56	53, 60	96	94, 97
LDL≥160	665	41	976	59	83	80, 86	75	72, 78	69	66, 73	95	93, 96
HDL<60	1,619	70	705	30	48	45, 50	84	81, 86	30	28, 33	97	95, 98
HDL<40	800	34	1,524	66	69	65, 72	78	76, 80	57	53, 60	96	95, 97
TG≥150	780	46	920	54	75	72, 78	76	73, 78	60	57, 64	95	94, 97
TG≥200	679	40	1,021	60	79	75, 82	73	71, 76	68	64, 71	95	93, 96
TC≥200	1,629	70	698	30	52	49, 54	94	92, 96	31	29, 34	99	99, 100
TC≥240	965	41	1,362	59	72	69, 75	86	84, 88	51	48, 54	98	97, 99

Abbreviations: CI, confidence interval; HDL, high-density lipoprotein cholesterol; LDL, low-density lipoprotein cholesterol; SN, sensitivity; SP, specificity; TC, total cholesterol; TG, triglycerides.

[a] The gold standard defines "condition present" as ≥2 distinct measurements meeting the threshold listed or ≥2 distinct lipid-controlling drug prescription fills and "condition absent" as 0 or 1 measurement meeting the threshold listed and 0 or 1 lipid-controlling drug prescription fill.

[b] Excluded from counts and from the calculation of sensitivity and specificity: N=40 participants who responded "don't know", N=686 with no available LDL gold standard, N=3 with no available HDL gold standard, and N=627 with no available TG gold standard. All participants have an available TC gold standard.

Table 3.3. Attenuated odds ratio for a hypercholesterolemia-disease association, Group Health, 1998-2007. Attenuated odds ratio when those who self-report untreated hypercholesterolemia are grouped with those who self-report treated hypercholesterolemia (AOR_1), and attenuated odds ratio when those who self-report untreated hypercholesterolemia are grouped with those who self-report no hypercholesterolemia (AOR_2). Assumes non-differential misclassification and the sensitivities and specificities reported in Table 3.2.

Gold standard (mg/dL)[a]	Prevalence (%)[b]	Attenuated odds ratio	OR with 100% sensitivity and 100% specificity				
			0.25	0.50	1.00	2.00	4.00
LDL≥130	52	AOR_1	0.46	0.67	1.00	1.48	2.04
		AOR_2	0.39	0.64	1.00	1.46	1.91
LDL≥160	41	AOR_1	0.53	0.70	1.00	1.49	2.20
		AOR_2	0.37	0.60	1.00	1.61	2.39
HDL<60	70	AOR_1	0.62	0.80	1.00	1.18	1.31
		AOR_2	0.53	0.76	1.00	1.20	1.35
HDL<40	34	AOR_1	0.62	0.76	1.00	1.38	1.89
		AOR_2	0.40	0.62	1.00	1.58	2.33
TG≥150	46	AOR_1	0.54	0.71	1.00	1.42	1.92
		AOR_2	0.40	0.63	1.00	1.51	2.08
TG≥200	40	AOR_1	0.57	0.73	1.00	1.43	2.01
		AOR_2	0.38	0.61	1.00	1.61	2.39
TC≥200	70	AOR_1	0.48	0.72	1.00	1.27	1.47
		AOR_2	0.49	0.73	1.00	1.23	1.39
TC≥240	41	AOR_1	0.48	0.67	1.00	1.50	2.15
		AOR_2	0.36	0.60	1.00	1.55	2.16

Abbreviations: AOR, attenuated odds ratio; CI, confidence interval; HDL, high-density lipoprotein cholesterol; LDL, low-density lipoprotein cholesterol; OR, odds ratio; TC, total cholesterol; TG, triglycerides.

[a] The gold standard defines "condition present" as ≥2 distinct measurements meeting the threshold listed or ≥2 distinct lipid-controlling drug prescription fills and "condition absent" as 0 or 1 measurement meeting the threshold listed and 0 or 1 lipid-controlling drug prescription fill.

[b] Calculation of AOR_1 and AOR_2 assumes the prevalence of hypercholesterolemia among controls is equal to the estimated prevalence for the given gold standard among all study participants.

Supplemental Table 3.1. Sensitivity and specificity of self-report of hypercholesterolemia from study questionnaire compared to gold standards based on clinical measurements and pharmacy records, stratified by the duration of available retrospective follow-up for gold standard, Group Health, 1998-2007

Duration of available retrospective follow-up, years[a]	Gold standard (mg/dL)[b]	Classification based on gold standard[c]				Classification of those who self-report untreated hypercholesterolemia							
		"Condition present"		"Condition absent"		"Condition present"				"Condition absent"			
						SN_1		SP_1		SN_2		SP_2	
		N	%	N	%	%	95% CI	%	95% CI	%	95% CI	%	95% CI
>0 to ≤10	LDL≥130	574	46	672	54	75	72, 79	81	78, 84	55	51, 59	96	94, 97
>10 to ≤20	LDL≥130	279	71	116	29	79	74, 84	89	82, 94	59	53, 65	98	94, 100
>0 to ≤10	LDL≥160	447	36	799	64	81	78, 85	75	72, 78	68	63, 72	95	93, 96
>10 to ≤20	LDL≥160	218	55	177	45	86	80, 90	73	66, 80	73	67, 79	96	92, 98
>0 to ≤10	HDL<60	294	55	245	45	52	46, 58	82	77, 87	35	30, 41	98	95, 99
>10 to ≤20	HDL<60	1,325	74	460	26	47	44, 49	85	81, 88	29	27, 32	97	95, 98
>0 to ≤10	HDL<40	153	28	386	72	69	61, 76	76	72, 81	60	52, 68	96	93, 97
>10 to ≤20	HDL<40	647	36	1,138	64	69	65, 72	78	76, 81	56	52, 60	96	95, 97
>0 to ≤10	TG≥150	447	38	727	62	73	69, 77	76	73, 79	60	55, 65	95	93, 97
>10 to ≤20	TG≥150	333	63	193	37	77	72, 82	75	68, 81	61	55, 66	97	93, 99
>0 to ≤10	TG≥200	390	33	784	67	77	72, 81	74	71, 77	67	62, 72	95	93, 96
>10 to ≤20	TG≥200	289	55	237	45	81	76, 86	70	64, 76	69	63, 74	96	93, 98
>0 to ≤10	TC≥200	291	55	241	45	59	53, 65	91	87, 95	36	30, 42	99	96, 100
>10 to ≤20	TC≥200	1,338	75	457	26	50	48, 53	96	94, 97	30	28, 33	100	99, 100
>0 to ≤10	TC≥240	174	32	358	67	76	69, 83	83	79, 87	57	49, 64	98	96, 99
>10 to ≤20	TC≥240	791	44	1,004	56	71	68, 74	87	85, 89	49	46, 53	98	97, 99

Abbreviations: CI, confidence interval; HDL, high-density lipoprotein cholesterol; LDL, low-density lipoprotein cholesterol; SN, sensitivity; SP, specificity; TC, total cholesterol;

TG, triglycerides.

[a] Duration between questionnaire and the earliest available measurement of that cholesterol trait prior to questionnaire or the earliest available lipid-controlling drug-prescription (whichever comes first).

[b] The gold standard defines "condition present" as ≥2 distinct measurements meeting the threshold listed or ≥2 distinct lipid-controlling drug prescription fills and "condition absent" as 0 or 1 measurement meeting the threshold listed and 0 or 1 lipid-controlling drug prescription fill.

[c] Excluded from counts and from the calculation of sensitivity and specificity: N=40 participants who responded "don't know", N=686 with no available LDL measurement, N=3 with no available HDL measurement, and N=627 with no available TG measurement. All participants have an available TC measurement.

Supplemental Table 3.2. Sensitivity and specificity of self-report of hypercholesterolemia from study questionnaire compared to gold standards based on clinical measurements and pharmacy records, stratified by year of questionnaire administration, Group Health, 1998-2007

						Classification of those who self-report untreated hypercholesterolemia							
		Classification based on gold standard[b]				"Condition present"				"Condition absent"			
Year of questionnaire administration	Gold standard (mg/dL)[a]	"Condition present"		"Condition absent"		SN_1		SP_1		SN_2		SP_2	
		N	%	N	%	%	95% CI	%	95% CI	%	95% CI	%	95% CI
1998-2003	LDL≥130	89	43	118	57	78	67, 86	84	76, 90	52	41, 62	94	88, 98
2004-2007	LDL≥130	764	53	670	47	76	73, 79	82	78, 85	57	53, 60	96	95, 98
1998-2003	LDL≥160	67	32	140	68	88	78, 95	79	72, 86	67	55, 78	94	89, 98
2004-2007	LDL≥160	598	42	836	58	82	79, 85	74	71, 77	70	66, 73	95	93, 96
1998-2003	HDL<60	269	66	141	34	37	31, 43	89	83, 94	19	14, 24	98	94, 100
2004-2007	HDL<60	1,350	71	564	29	50	47, 52	82	79, 85	33	30, 35	97	95, 98
1998-2003	HDL<40	113	28	297	72	58	48, 67	84	79, 88	42	32, 51	98	96, 99
2004-2007	HDL<40	687	36	1,227	64	70	67, 74	77	74, 79	59	56, 63	96	94, 97
1998-2003	TG≥150	102	44	132	56	73	63, 81	85	78, 90	48	38, 58	97	92, 99
2004-2007	TG≥150	678	46	788	54	75	72, 78	74	71, 77	62	58, 66	95	93, 97
1998-2003	TG≥200	78	33	156	67	81	70, 89	80	73, 86	60	49, 71	96	92, 99
2004-2007	TG≥200	601	41	865	59	78	75, 82	72	69, 75	69	65, 73	95	93, 96
1998-2003	TC≥200	259	63	152	37	42	36, 48	96	92, 99	20	15, 25	99	96, 100
2004-2007	TC≥200	1,370	72	546	29	54	51, 56	94	91, 96	33	31, 36	99	98, 100
1998-2003	TC≥240	135	33	276	67	64	56, 72	90	86, 93	37	29, 46	99	97, 100
2004-2007	TC≥240	830	43	1,086	57	73	70, 76	85	83, 87	53	50, 56	98	97, 99

Abbreviations: CI, confidence interval; HDL, high-density lipoprotein cholesterol; LDL, low-density lipoprotein cholesterol; SN, sensitivity; SP, specificity; TC, total cholesterol; TG, triglycerides.

[a] The gold standard defines "condition present" as ≥2 distinct measurements meeting the threshold listed or ≥2 distinct lipid-controlling drug prescription fills and "condition absent" as 0 or 1 measurement meeting the threshold listed and 0 or 1 lipid-controlling drug prescription fill.

[b] Excluded from counts and from the calculation of sensitivity and specificity: N=40 participants who responded "don't know", N=686 with no available LDL measurement, N=3 with no available HDL measurement, and N=627 with no available TG measurement. All participants have an available TC gold measurement.

4. Blood lipids and colorectal polyps: testing an etiologic hypothesis using phenotypic measurements and Mendelian randomization

ABSTRACT

Studies linking cholesterol levels to the development of colorectal neoplasia are inconsistent. Recent genetic studies have identified >100 loci associated with lipids. We genotyped individuals who received a colonoscopy at Group Health (1998-2007), and had available pre-colonoscopy lipid measurements, for 96 of 102 single-nucleotide polymorphisms (SNPs) identified by the Global Lipids Genetics Consortium. Participants included 518 non-advanced adenoma cases, 139 advanced adenoma cases, 380 non-adenomatous polyp cases, and 754 polyp-free controls. Advanced adenoma cases were more likely than controls to have higher pre-colonoscopy zenith low-density lipoprotein (LDL), triglycerides (TG), and total cholesterol (TC) (odds ratio, OR, per 20 mg/dL LDL increase: 1.16, 95% confidence interval, CI, 1.03-1.30; per 40 mg/dL TG increase: 1.09, 1.03-1.16; and per 20 mg/dL TC increase: 1.09, 1.02-1.18). Genotype-polyp ORs using allele scores were not statistically significant (OR per increase in score scaled to a 20 mg/dL LDL increase: 1.17, 0.78-1.75; a 40 mg/dL TG increase: 1.12, 0.91-1.38; a 20 mg/dL TC increase: 0.99, 0.71-1.38). SNPs with the largest magnitude association with lipids were not associated with polyps, and SNPs associated with polyps may function through alternative pathways. There was insufficient evidence to conclude that genetic susceptibility to dyslipidemia is associated with colorectal polyps.

It is unclear whether primary and secondary dyslipidemia are risk factors for colorectal neoplasia. Many epidemiological studies have observed an increased prevalence of adenomatous polyps (adenomas) among those with cholesterol profiles associated with unfavorable cardiovascular disease outcomes (2, 3, 9, 11, 84, 147, 148). This evidence is consistent with autopsy studies that suggest a tendency for those with colorectal polyps to also have atherosclerosis (149), and biological studies of the neoplastic potential of bile acids and other lipid metabolites (60, 150). On the other hand, several studies of invasive colorectal cancer have reported an *inverse* association with cholesterol (151). Although often explained by reverse causation, as cholesterol measured shortly after diagnosis may reflect the cholesterol-lowering effects of chemotherapy (12), a decrease in cholesterol levels around 2 years prior to cancer diagnosis, possibly due to preclinical disease, cannot be ruled out (78). Randomized controlled trials of lipid-controlling drugs have not observed an increased or decreased risk of gastrointestinal malignancies (65, 152), and observational studies of lipid-contolling drug use focusing on colorectal polyps have been inconsistent (67-69).

It has been challenging for observational studies to determine that the association between colorectal neoplasia and dyslipidemia is unconfounded by shared risk factors including high-fat diet, obesity, insulin resistance, smoking, and sedentary lifestyle. Mendelian randomization has been suggested as a potential solution (28). Under assumptions employed in instrumental variables analysis (153), genotypes can serve as proxies for phenotypes so that genotype-disease associations mimic phenotype-disease associations, but with limited potential for bias from confounding and reverse causation (30). In this manner, Mendelian randomization studies use the distribution of alleles in the population to simulate randomized assignment to lower or higher cholesterol over the life course.

In the case of polygenic dyslipidemia, which is common (154) and highly hereditable (155), the Global Lipids Genetics Consortium (GLGC) genome-wide association study (GWAS) identified 102 germline single-nucleotide polymorphisms (SNPs) across 95 genes reaching genome-wide statistical significance for associations with blood concentrations of low-density lipoprotein cholesterol (LDL), high-density lipoprotein cholesterol (HDL), triglycerides (TG), or total cholesterol (TC) (41). Using these SNPs, we conducted a Mendelian randomization study in a sample of men and women who underwent colonoscopy at Group Health, a large healthcare system in Washington State.

MATERIALS AND METHODS

Study population

Participants, ages 25-79, were enrollees of Group Health who received a colonoscopy for any indication from 1998-2007 (49). Only those with at least 3 years of enrollment with Group Health and no previous colonoscopy within 1 year prior to the study colonoscopy were eligible. Individuals with a personal history of ulcerative colitis, Crohn's disease, or invasive colorectal cancer were ineligible, and those diagnosed with these conditions during the study colonoscopy were excluded. All participants completed a health history interview and provided informed consent to access biospecimens and medical records. Study protocols were approved by the Institutional Review Boards of Group Health Research Institute and the Fred Hutchinson Cancer Research Center.

Outcome ascertainment

Diagnostic colorectal biopsies collected during colonoscopy were reviewed by 2 study pathologists. Adenomas were distinguished from non-adenomatous polyps, which included hyperplastic polyps, traditional serrated adenomas, and sessile serrated adenomas. An advanced adenoma was defined as any tubular, tubulovillous, or villous adenoma ≥10 mm in diameter, with ≥20% villous components, or high-grade dysplasia (58). Participants were classified into 4 groups: 1) non-advanced adenomas cases; 2) advanced adenoma cases; 3) non-adenomatous polyp cases; and 4) those for whom colonoscopy revealed no polyps (controls). Participants with synchronous polyps of different types were grouped based on the lesion(s) suspected to have the highest malignant potential (advanced adenoma prioritized above the other 2 types, and non-advanced adenoma prioritized above non-adenomatous polyps).

Phenotype measurement

Group Health maintains a database of laboratory test results dating back to the 1980s. All available LDL, HDL, TG, and TC measurements were extracted from at most 20 years prior to each participant's study colonoscopy. We determined each participant's highest pre-colonoscopy LDL, highest pre-colonoscopy TG, highest pre-colonoscopy TC measurement (zenith), and lowest pre-colonoscopy HDL measurement (nadir). If 1 of 4 was missing but the other 3 were measured on the same day, which occurred for fewer than 2% of measurement occasions, the Friedwald equation was used to calculate the missing value provided TG values did not exceed 400 mg/dL (124). LDL or TG measurements were unavailable for about 30% of participants, as these were not routinely used to assess cardiovascular disease risk at Group Health until the later

period of data collection. Measurements were intended to be fasting, but compliance with this requirement could not be verified.

Information on lipid-controlling drug prescriptions dispensed at eligible pharmacies, including generic name, dose, and fill-date, were extracted from electronic pharmacy records at Group Health (57). Medication types included both 3-hydroxy-3-methylglutaryl coenzyme A reductase inhibitors, commonly referred to as statins (atorvastatin, fluvastatin, lovastatin, pravastatin, rosuvastatin, simvastatin), as well as prescription non-statins (cholestyramine, colesevelam, colestipol, ezetimbe, fenofibrate, gemfibrozil, niacin).

Genotype measurement

Participants were asked to provide a blood or buccal sample for genetic analysis (56). Genotyping was performed using a custom GoldenGate assay from Illumina (San Diego, CA). Prior to genotyping, 11 of the 102 SNPs identified by the GLGC (41) were projected to have low likelihood of success based on Illumina's Assay Design Tool. For 5 of these SNPs (rs7515577, *EVI5*; rs1042034, *APOB*; rs9488822, *FRK*; rs12967135, *MC4R*; and rs7255436; *ANGPTL4*, we identified proxies based on linkage disequilibrium estimates from HapMap CEUs (European Caucasians). Six SNPs for which no suitable proxy could be identified were excluded (rs1367117, *APOB*; rs13238203, *TYW1B*; rs4759375, *SBNO1*; rs2652834, *LACTB*; rs7241918, *LIPG*; rs2277862, *ERGIC3*). For quality control purposes, 77 samples were genotyped as replicates, along with 30 parent-child trios from the CEU collection. All SNPs were tested for departures from Hardy-Weinberg equilibrium.

Study exclusions

A total of 2,506 participants completed the interview (1,037 controls, 700 non-advanced adenoma cases, 175 advanced adenoma cases, and 594 non-adenomatous polyp cases). DNA was unavailable from 250 (24%) controls, 158 (23%) non-advanced adenoma cases, 34 (19%) advanced adenoma cases, and 198 (33%) non-adenomatous polyp cases. Nineteen had insufficient DNA for genotyping (9 controls, 6 non-advanced adenoma cases, and 4 non-adenomatous polyp cases). A further 56 had no available lipid measurements within 20 years prior to the study colonoscopy (24 controls, 18 non-advanced adenoma cases, 2 advanced adenoma cases, and 12 non-adenomatous polyp cases).

Statistical analyses

For each of LDL, HDL, TG, and TC, we estimated 3 primary associations: 1) lipid-polyp odds ratios (OR) with 95% confidence intervals (CI) comparing each case group to controls using polytomous logistic regression; 2) genotype-lipid associations using single-SNP and multi-SNP ordinary linear regression; and 3) genotype-polyp Mendelian randomization ORs using 2-stage linear-logistic regression (156).

The 2-stage Mendelian randomization estimator can be expressed as $\hat{\beta}_{\text{lipid-polyp}} = \frac{\hat{\beta}_{\text{genotype-polyp}}}{\hat{\beta}_{\text{genotype-lipid}}}$. Because slopes from logistic regression parameterize log-ORs, we have $\widehat{OR}_{\text{lipid-polyp}} = e^{\frac{\widehat{log\text{-}OR}_{\text{genotype-polyp}}}{\hat{\beta}_{\text{genotype-lipid}}}} = \widehat{OR}_{\text{genotype-polyp}}^{\frac{1}{\hat{\beta}_{\text{genotype-lipid}}}}$ (157). To improve interpretability, we chose to scale lipid-polyp associations per 20 mg/dL increase in zenith LDL and zenith TC, per 10 mg/dL decrease in nadir HDL, and per 40 mg/dL increase in zenith TG. The 2-stage

Mendelian randomization ORs were similarly scaled. That is, lipid-polyp associations for a Δ-unit increase in lipid phenotype are reported as $\widehat{OR}_{\text{genotype-polyp}}^{\frac{\Delta}{\hat{\beta}_{\text{genotype-lipid}}}}$.

In the GLGC GWAS, several SNPs achieved genome-wide statistical significance ($P<5\times10^{-8}$) for 2 or more lipid traits. Of the 96 total SNPs we genotyped, there were 36 identified for LDL, 44 for HDL, 31 for TG, and 49 for TC. We performed single-SNP and 2 types of multi-SNP analyses: 1) using an allele score for each lipid trait created by counting alleles associated with an increased mean of that trait in the GLGC GWAS, weighted by effect size from their analysis (for HDL, the score was based on alleles associated with decreased mean HDL) (158); and 2) without using an allele score, we regressed per-allele log-ORs on the association from the GLGC GWAS and tested for the statistical significance of slopes using inverse-variance-weighted linear regression (159). When calculating allele scores that included SNPs with missing genotypes, we imputed the case- or control-group-specific mean score calculated among participants not missing any genotypes. The appropriateness of assuming a linear association with the trait-specific allele score was assessed using a categorical allele-score variable based on deciles in controls.

We evaluated both minimally- and fully-adjusted models for all estimates. The former included age at colonoscopy (<50, 50-60, 60-70, ≥70 years), sex (male, female), race (White/Caucasian, Black/African American, Asian/Pacific Islander, other), and data-collection period (1998-2003, 2004-2007), and the second model further adjusted for highest level of education (high school or less, some college, college, graduate), body mass index (BMI; <25, 25-30, ≥30 kg/m^2), family history of colorectal cancer in first-degree relatives (no, yes), personal history of diabetes mellitus (no, yes), nonsteroidal anti-inflammatory drug (NSAID) use (never,

former, current), estrogen-alone use (only for women; no, yes), estrogen-plus-progestin use (only for women; no, yes), cigarette smoking (never, former, current), alcohol consumption (0, >0-<7,7-14, ≥14 drinks/week), vegetable consumption (0-<1, 1-2, 2-3, ≥3 servings/day), fruit consumption (0-<1, 1-2, 2-3, ≥3 servings/day), recreational and exercise physical activity (0, >0-<1, 1-2, 2-6, ≥6 hours/week), and prior lower endoscopy (earlier than approximately 2 years before the study colonoscopy; no, yes).

F-statistics, which have been traditionally used to assess instrument strength in instrumental variables analyses, are provided for genotype-lipid associations (160). Because SNPs were selected from previous GWAS results, and the overall hypotheses were based on a single multi-SNP summary measure per trait, we did not penalize for multiple comparisons except to account for the fact that analyses were repeated for 4 different lipid traits. Thus, we considered 2-sided *P*-values$\leq 0.05/4 \approx 0.01$ to denote statistical significance. We also report single-SNP associations with $P \leq 0.01$, but acknowledge that this does not control for the family-wise error at $\alpha=0.01$. All analyses were performed using SAS 9.2 (Cary, NC) or R 3.0.0 (Vienna, Austria).

Sensitivity analyses

For some participants using lipid-controlling drugs, only post-treatment blood lipid values may be available from Group Health records. In this case, the true zenith LDL may be higher than observed (161). To account for this, we considered 3 sensitivity analyses: 1) excluding all lipid-controlling drug users (N=503); 2) excluding lipid-controlling drug users with pre-colonoscopy zenith LDL<130 mg/dL, borderline high LDL as defined by the Adult Treatment Panel III (125) (N=90); and 3) including users with pre-colonoscopy zenith LDL<130

mg/dL, but with the zenith LDL imputed as the maximum of 100 mg/dL or 30 mg/dL higher than the observed zenith LDL. We selected 30 mg/dL as a conservative estimate for the mean 1-year treatment effect of statins from randomized controlled trials (44).

RESULTS

Participant characteristics

Most participants were White/Caucasian, and the mean age was 60 years (Table 4.1). The majority of those with adenomas were men, whereas most polyp-free controls were women. Adenoma cases were slightly more likely than controls to have used statins prior to colonoscopy; few (2%) had prescriptions for only non-statin lipid-controlling drugs.

Lipid-polyp associations

Zenith LDL, zenith TG, and zenith TC were each associated with increased odds of advanced adenomas. There was also evidence that zenith LDL, but not other lipid traits, was associated with increased odds of non-advanced adenomas and non-adenomatous polyps (Table 2). Zenith LDL was also associated with increased odds of polyps among participants with no evidence of lipid-controlling drug prescriptions, and in analyses that attempted to correct for potentially unobserved pre-colonoscopy extremes in LDL (Supplemental Table 4.1).

Genotyping results

Genotyping quality was excellent; replicate samples were genotyped with >99% concordance. Most of the 96 genotyped SNPs were missing for <1% of participants. Exceptions were rs2068888 (*CYP26A1*) missing for 74% of participants, and rs7134375, (*PDE3A*) and

rs4420638 (*APOE/APOC1*) both missing for 26% of participants. For all but 1 SNP, the minor allele observed among our controls matched the minor allele reported from the GLGC GWAS. For rs4129767 (*PGS1*) the G-allele frequency was 54% in controls, but was 49% in the GLGC GWAS, thus we report associations per G allele of this SNP to be consistent with the GLGC GWAS.

Genotype-lipid associations

Although we assessed SNPs that met genome-wide significance in the GLGC GWAS (Web Table 2), with our smaller sample size, only 11 SNPs were associated with their respective phenotype among controls at $P \leq 0.01$ (Supplemental Tables 4.3-4.6). All 11 were in the same direction, but somewhat stronger, than associations reported by the GLGC GWAS. Genotype-lipid associations using GLGC-weighted allele scores were highly statistically significant, with *P*-values ranging from 1×10^{-6} for the LDL allele score to 1×10^{-17} for the HDL allele score in minimally-adjusted models among controls (Supplemental Table 4.7). Allele scores were not associated with other covariates including BMI, smoking, or fruit and vegetable consumption (Supplemental Table 4.8).

Single-SNP genotype-polyp associations

Of the 11 SNPs associated with lipid phenotypes in controls at $P \leq 0.01$, only 1 SNP (rs12670798 of *DNAH11*) was among the 7 SNPs associated with any type of colorectal polyp at $P \leq 0.01$ (Figure 4.1). Alleles of this SNP, associated with increased mean LDL and TC in the GLGC GWAS, were associated with increased prevalence of advanced adenomas.

SNPs of the *APOE/APOC1* locus were associated with both lipid phenotypes in controls and polyps, but by different uncorrelated SNPs (rs439401 and rs4420638). For both SNPs, the

allele associated with lipid phenotypes that convey increased cardiovascular disease risk (increased LDL and decreased HDL for rs4420638 and increased TG for rs439401) was associated with increased odds of non-advanced adenomas. Advanced adenoma cases were more likely than controls to carry alleles of a SNP of *PLEC* (rs11136341) associated with increased LDL. Two loci associated with increased TG (the major allele of a SNP of *AFF1*; rs442177, and the minor allele of a SNP of *NAT2*; rs1495741) were associated with lower odds of non-advanced adenomas. The minor allele SNP of *MC4R* (rs10871777) was associated with non-adenomatous polyps, and that of the *PSKH1/LCAT* locus (rs16942887) was associated with advanced adenomas, but these alleles were associated with HDL in opposite directions in the GLGC GWAS.

Multi-SNP genotype-polyp associations

Mendelian randomization estimates for genotype-polyp associations based on allele scores were less precise than estimates from lipid-polyp analyses. None was statistically significant (Table 4.3). The linearity of the genotype-lipid association based on deciles of the allele score in controls is displayed in Figure 4.2, which also shows no discernible trends in the associations with polyps even at the extremes of the allele-score distributions. Multi-SNP analyses without using allele scores (Figure 4.3) or accounting for lipid-controlling drug use (Supplemental Table 4.9) also revealed no statistically significant associations.

DISCUSSION

We found that larger extremes in LDL, TG, and TC occurring, on average, about 4 years before colonoscopy, were associated with the prevalence of advanced adenomas, those lesions

most likely to progress to invasive colorectal cancer (58). Although the directions of associations from allele scores support this observation for LDL and TG, the evidence was not strong, particularly in light of the apparent inconsistency between which SNPs were associated with lipid phenotypes and which were associated with polyps. In general, the polymorphisms with the largest magnitude per-allele associations with lipid phenotypes were not associated with colorectal polyps.

The Mendelian randomization approach was first motivated by epidemiologic studies of cholesterol and cancer nearly 30 years ago (162). Pre-GWAS Mendelian randomization studies have evaluated candidate gene variants with respect to cancer of any type, and consistent with longitudinal evidence, concluded that inverse lipid-cancer associations likely result from reverse causation (163, 164). Our study is the first Mendelian randomization analysis to target pre-diagnostic cholesterol and precursors for colorectal cancer, aggregating information across >30 GWAS-identified variants per trait.

Estimating phenotype-disease associations from genotypes requires strong assumptions (153). Alleles must function to alter blood lipid levels without unmeasured common causes of both the polymorphism and polyp occurrence, and without the alleles being involved in mechanisms that influence colorectal polyp formation separate from the mechanisms by which they alter blood lipid levels (i.e., no genetic pleiotropy) (165). Mendelian randomization analyses of traits with complex biology are difficult to interpret. Some of the SNPs we evaluated may be inappropriate for use as instrumental variables due to pleiotropy or weak-instrument bias (166, 167).

Consistent with previous studies, we observed associations with polyps for variants of genes such as *APOE/APOC1*, *NAT2*, and *MC4R*. The *APOE*-e4 allele has been found to be associated with reduced risk of proximal adenoma (35, 39, 168). The 2 SNPs near *APOE/APOC1* identified by the GLGC GWAS (rs43940 and rs4420638), however, have low correlation with SNPs that define the e1-e4 alleles (rs7412 and rs429358). With respect to invasive colorectal cancer, reported associations with *APOE* genotype have been limited to specific subgroups not supported by prior hypotheses (37, 169).

Although we observed an association between polyps and a SNP of *NAT2*, this relation may not be mediated by the gene's influence on lipid levels. *NAT2* functions to activate potentially carcinogenic heterocyclic amines, and has been extensively studied in relation to adenoma and invasive colorectal cancer risk (170), particularly involving interactions with red meat consumption (171) and smoking (172). The allele of rs1495741 associated with reduced odds of colorectal lesions in our study, is a known marker of the rapid acetylator phenotype (173) and has been associated with reduced risk of other cancers (174).

Study participants with non-adenomatous polyps were more likely than controls to carry alleles of a SNP of *MC4R* known to be related to lower HDL. *MC4R* encodes a receptor that interacts with α-melanocyte-stimulating hormone to regulate appetite, energy balance, and obesity, and is known to harbor SNPs (highly correlated with the HDL SNP rs12967135) related to BMI, waist circumference, and insulin resistance in GWAS (175, 176). Common SNPs near *MC4R* were among the top-hits in the first stage of a GWAS for invasive colorectal cancer, but did not achieve statistical significance when combined with a replication sample (177). The association between BMI and adenoma is well-established (178) and this variant of *MC4R* may be a marker of obesity that is independent of cholesterol.

Although replication is needed, we identified 4 other lipid-related loci associated with colorectal polyps, *AFF1* (rs442177), *PLEC* (rs11136341), *PSKH1/LCAT* (rs16942887), and *DNAH11* (rs12670798). These loci do not appear to be previously related to colorectal neoplasia, although variants of *DNAH11* have been related to the risk of at least one type of malignancy (179). Associations between colorectal cancer risk and other candidate SNPs related to cholesterol metabolism have been reported (168, 180), including a SNP of *HMGCR* (rs12654264) that modified the association between statin use and colorectal cancer risk (181). This variant is highly correlated with a SNP in the same gene identified by the GLGC GWAS (rs12916) for which we found no evidence of an association with polyps.

The ability to compare estimates from Mendelian randomization to those from clinical lipids measurements was a key motivation for collecting data on both genotype and phenotype. This also permitted an internal assessment of the strength of instrumental variables. It is not typically necessary to adjust for disease risk factors in genetic association studies, and the minimally-adjusted estimates we report align closely with the classical notions of Mendelian randomization. We did, however, utilize detailed information on multiple characteristics associated with the risk of colorectal neoplasia ascertained from interview to calculate fully-adjusted estimates as an attempt to isolate associations independent of pleiotropic pathways (182).

Previous colonoscopy studies of cholesterol exposures often measure blood lipids on the day of exam, which may not reflect the most etiologically-relevant exposure period. Our use of extreme values from clinical measurements over several years prior to colonoscopy, although subject to outliers, likely provides a better marker of dyslipidemia status than can be captured from a measurement at the time of colonoscopy. The sensitivity of using the extreme values is

underscored by the fact that a GWAS-identified variant of *CETP* (rs3764261) met genome-wide statistical significance for the association with nadir HDL in our relatively modest sample of 754 controls.

Genes in cholesterol metabolism pathways have not been among the loci reaching genome-wide statistical significance from GWAS of colorectal polyps (183, 184) or invasive colorectal cancer (185). Thus, we used trait-specific allele scores in an attempt to operationalize sufficiently strong instrumental variables. In the time since data collection for our study was completed, the GLGC reported an additional 62 SNPs (157 total) associated with LDL, HDL, TG, and TC at $P<5\times10^{-8}$ (186), but it is unclear if inclusion of additional SNPs will strengthen the instrument.

A primary limitation of our study is the sample size. Achieving 80% statistical power to detect the magnitude of lipid-polyp associations we observed would have required allele scores to explain >95% of the variance in extreme lipid levels (187). This is not the case, as it is estimated that the 102 SNPs from the GLGC GWAS collectively explain approximately 12% of total variation, or about 30% of the expected genetic variation, in each lipid trait (41). Sample size considerations prevented stratification by additional factors such as anatomic location of lesions. Larger consortium-based studies will benefit from enhanced statistical power, but may have limited ability to harmonize pathology information, past lipid trajectories, and pharmacy data.

In summary, our data do not support the conclusion that common genetic variation controlling cholesterol levels is involved in polyp formation. Observed associations between blood lipid concentrations and colorectal polyps may be non-causal. For instance, dyslipidemia

may be a marker of the type of visceral adiposity and dietary exposures that promote neoplastic growth in the colon and rectum.

Table 4.1. Characteristics of colorectal polyp cases and controls, Group Health, 1998-2007

Characteristics[a]	Controls (N=754)	Non-advanced adenomas[b] (N=518)	Advanced adenomas[c] (N=139)	Non-adenomatous polyps[d] (N=380)
At colonoscopy				
Age at colonoscopy (years), N (%)				
<50	65 (9)	28 (5)	5 (4)	19 (5)
50-59	291 (39)	192 (37)	53 (38)	177 (47)
60-69	272 (36)	194 (37)	51 (37)	133 (35)
70-79	126 (17)	104 (20)	30 (22)	51 (13)
White/Caucasian, N (%)	656 (87)	439 (85)	121 (87)	337 (89)
Male, N (%)	306 (41)	277 (53)	78 (56)	170 (45)
Prior to colonoscopy				
Statin use, N (%)	194 (26)	162 (31)	41 (30)	97 (26)
Years between first statin prescription and colonoscopy, mean (95% CI)	4.2 (3.7, 4.7)	4.3 (3.8, 4.9)	4.8 (3.6, 6.0)	4.1 (3.3, 4.8)
LDL				
LDL measurement(s) available, N (%)	513 (68)	390 (75)	103 (74)	269 (71)
Zenith LDL (mg/dL), mean (95% CI)	143 (140, 147)	149 (146, 153)	153 (145, 160)	148 (144, 153)
Years between zenith LDL and colonoscopy, mean (95% CI)	4.9 (4.5, 5.4)	4.5 (4.0, 4.9)	5.3 (4.3, 6.2)	4.9 (4.3, 5.5)
HDL				
HDL measurement(s) available, N (%)	752 (100)	518 (100)	139 (100)	380 (100)
Nadir HDL (mg/dL), mean (95% CI)	49 (48, 50)	47 (46, 49)	48 (45, 50)	49 (48, 51)
Years between nadir HDL and colonoscopy, mean (95% CI)	9.4 (9.0, 9.7)	9.2 (8.8, 9.6)	8.4 (7.5, 9.3)	9.7 (9.2, 10.2)
TG				
TG measurement(s) available, N (%)	535 (71)	402 (78)	105 (76)	274 (65)
Zenith TG (mg/dL), mean (95% CI)	190 (179, 201)	208 (194, 222)	240 (208, 272)	186 (170, 202)
Years between zenith TG and colonoscopy, mean (95% CI)	4.8 (4.4, 5.2)	4.4 (4.0, 4.9)	5.0 (4.2, 5.9)	4.7 (4.2, 5.3)
TC				
TC measurement(s) available, N	754 (100)	518 (100)	139 (100)	380 (100)

(%)				
Zenith TC (mg/dL), mean (95% CI)	239 (236, 242)	243 (239, 248)	252 (242, 262)	241 (237, 246)
Years between zenith TC and colonoscopy, mean (95% CI)	6.6 (6.3, 7.0)	6.7 (6.3, 7.2)	6.7 (5.8, 7.6)	6.3 (5.7, 6.8)

[a] Means and percentages are unadjusted.

[b] Non-advanced adenoma cases have at least one tubular or tubulovillous adenoma, all <10 mm in diameter, with <20% villous components, and no high-grade dysplasia.

[c] Adenoma cases have at least one tubular, tubulovillous, or villous adenoma ≥10 mm in diameter, with ≥20% villous components, or high-grade dysplasia.

[d] Non-adenomatous polyp cases had at least one hyperplastic polyp, traditional serrated adenoma, or sessile serrated adenoma, and no adenomas.

Abbreviations: CI, confidence interval; HDL, high-density lipoprotein cholesterol; LDL, low-density lipoprotein cholesterol; TC, total cholesterol; TG, triglycerides.

Table 4.2. Lipid-polyp associations using the highest available pre-colonoscopy blood lipid measurement (lowest for HDL), Group Health, 1998-2007

			Non-advanced adenomas (N=518) vs. controls (N=754)		Advanced adenomas (N=139) vs. controls (N=754)		Non-adenomatous polyps (N=380) vs. controls (N=754)	
Phenotype[a]	Adjustment[b]	Unit change for OR	OR (95% CI)	*P*	OR (95% CI)	*P*	OR (95% CI)	*P*
Zenith LDL	Minimal	+20 mg/dL	1.06 (0.99, 1.14)	0.07	1.11 (1.00, 1.24)	0.05	1.06 (0.98, 1.14)	0.14
Zenith LDL	Full	+20 mg/dL	1.07 (0.99, 1.15)	0.08	1.16 (1.03, 1.30)	0.01	1.10 (1.01, 1.20)	0.03
Nadir HDL	Minimal	-10 mg/dL	1.02 (0.93, 1.11)	0.74	0.97 (0.85, 1.12)	0.71	0.97 (0.88, 1.07)	0.51
Nadir HDL	Full	-10 mg/dL	0.94 (0.86, 1.05)	0.28	0.93 (0.79, 1.09)	0.37	0.98 (0.88, 1.10)	0.72
Zenith TG	Minimal	+40 mg/dL	1.03 (0.99, 1.07)	0.12	1.08 (1.03, 1.14)	0.004	1.00 (0.95, 1.05)	1.00
Zenith TG	Full	+40 mg/dL	1.03 (0.98, 1.08)	0.21	1.09 (1.03, 1.16)	0.006	1.02 (0.96, 1.07)	0.53
Zenith TC	Minimal	+20 mg/dL	1.03 (0.98, 1.08)	0.31	1.09 (1.02, 1.17)	0.01	1.02 (0.97, 1.08)	0.50
Zenith TC	Full	+20 mg/dL	1.03 (0.98, 1.09)	0.29	1.09 (1.02, 1.18)	0.02	1.05 (0.99, 1.11)	0.12

[a] Polytomous logistic regression model with case-control status as the dependent variable and the highest lipid measurement (lowest for HDL) in 20 years prior to colonoscopy as the independent variable.

[b] Minimally-adjusted OR is adjusted for age at colonoscopy, sex, race, and data-collection period. Fully-adjusted OR is adjusted for age at colonoscopy, sex, race, education, BMI, NSAID use, family history of CRC, estrogen-only use (in women), estrogen-plus-progestin use (in women), cigarette smoking, alcohol consumption, diabetes mellitus, fruit servings per day, vegetable servings per day, recreational and exercise physical activity, prior endoscopy (approximately 2 years before study colonoscopy), and data-collection period.

Abbreviations: BMI, body mass index; CI, confidence interval; HDL, high-density lipoprotein cholesterol; LDL, low-density lipoprotein cholesterol; NSAID, nonsteroidal anti-inflammatory drug; OR, odds ratio; TC, total cholesterol; TG, triglycerides.

Table 4.3. Genotype-polyp associations using trait-specific multi-SNP allele scores, Group Health, 1998-2007

Genetic exposure[a]	Adjustment[b]	Unit change for OR[c]	Non-advanced adenomas (N=518) vs. controls (N=754) OR (95% CI)	*P*	Advanced adenomas (N=139) vs. controls (N=754) OR (95% CI)	*P*	Non-adenomatous polyps (N=380) vs. controls (N=754) OR (95% CI)	*P*
LDL Allele Score	Minimal	+8.7 score	0.99 (0.77, 1.27)	0.95	1.17 (0.78, 1.75)	0.46	1.13 (0.86, 1.49)	0.40
LDL Allele Score	Full	+8.3 score	1.03 (0.80, 1.32)	0.84	1.11 (0.74, 1.66)	0.63	1.15 (0.87, 1.52)	0.32
HDL Allele Score	Minimal	+11.1 score	1.05 (0.79, 1.39)	0.76	0.95 (0.60, 1.48)	0.81	1.03 (0.75, 1.41)	0.87
HDL Allele Score	Full	+11.1 score	1.01 (0.75, 1.36)	0.96	1.01 (0.63, 1.64)	0.96	1.05 (0.75, 1.47)	0.77
TG Allele Score	Minimal	+4.1 score	1.09 (0.95, 1.24)	0.21	1.12 (0.91, 1.38)	0.30	1.07 (0.92, 1.23)	0.39
TG Allele Score	Full	+4.0 score	1.11 (0.97, 1.27)	0.13	1.15 (0.93, 1.43)	0.19	1.11 (0.96, 1.29)	0.17
TC Allele Score	Minimal	+8.0 score	1.02 (0.83, 1.26)	0.84	0.99 (0.71, 1.38)	0.94	1.08 (0.86, 1.37)	0.51
TC Allele Score	Full	+8.3 score	1.08 (0.85, 1.36)	0.53	0.98 (0.68, 1.41)	0.89	1.12 (0.87, 1.45)	0.38

[a] Polytomous logistic regression model with case-control status as the dependent variable and count of alleles associated with increasing lipid measurement (decreasing for HDL) in the GLGC GWAS weighted by the effect size from GLGC GWAS as the independent variable.

[b] Minimal OR adjusted for age at colonoscopy, sex, race, and data-collection period. Full OR adjusted for age at colonoscopy, sex, race, education, BMI, NSAID use, family history of CRC, estrogen-only use (in women), estrogen-plus-progestin use (in women), cigarette smoking, alcohol consumption, diabetes mellitus, fruit servings per day, vegetable servings per day, recreational and exercise physical activity, prior endoscopy (approximately 2 years before study colonoscopy), and data-collection period.

[c] The unit change in allele score is based on estimated slope in controls, scaled to correspond to a 20 mg/dL increase in highest LDL prior to colonoscopy, a 10 mg/dL decrease in lowest HDL prior to colonoscopy, a 40 mg/dL increase in highest TG prior to colonoscopy, or a 20 mg/dL increase in highest TC prior to colonoscopy.

Abbreviations: BMI, body mass index; CI, confidence interval; GLGC, Global Lipids Genetic Consortium; GWAS, genome-wide

association study; HDL, high-density lipoprotein cholesterol; LDL, low-density lipoprotein cholesterol; NSAID, nonsteroidal anti-inflammatory drug; OR, odds ratio; SNP, single-nucleotide polymorphism; TC, total cholesterol; TG, triglycerides.

GLGC GWAS results are from Teslovich TM, Musunuru K, Smith AV, et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* 2010;466(7307):707-13.

Figure 4.1. Two-way Manhattan plot of the *P*-values from single-SNP associations with lipid-component specific phenotype among controls (X-axis) and case-control status (Y-axis) among colonoscopy recipients, Group Health,1998-2007. All SNPs reached genome-wide statistical significance in the GLGC GWAS. SNPs with statistically significant associations with both lipid phenotypes among controls in our data and polyps should cluster in the upper right area of the plot. Regression models are adjusted for age, sex, race, and data-collection period. Selected SNPs are labelled by gene name. GLGC, Global Lipids Genetic Consortium; GWAS, genome-wide association study; HDL, high-density lipoprotein cholesterol; LDL, low-density lipoprotein cholesterol; SNP, single-nucleotide polymorphism; TC, total cholesterol; TG triglycerides.

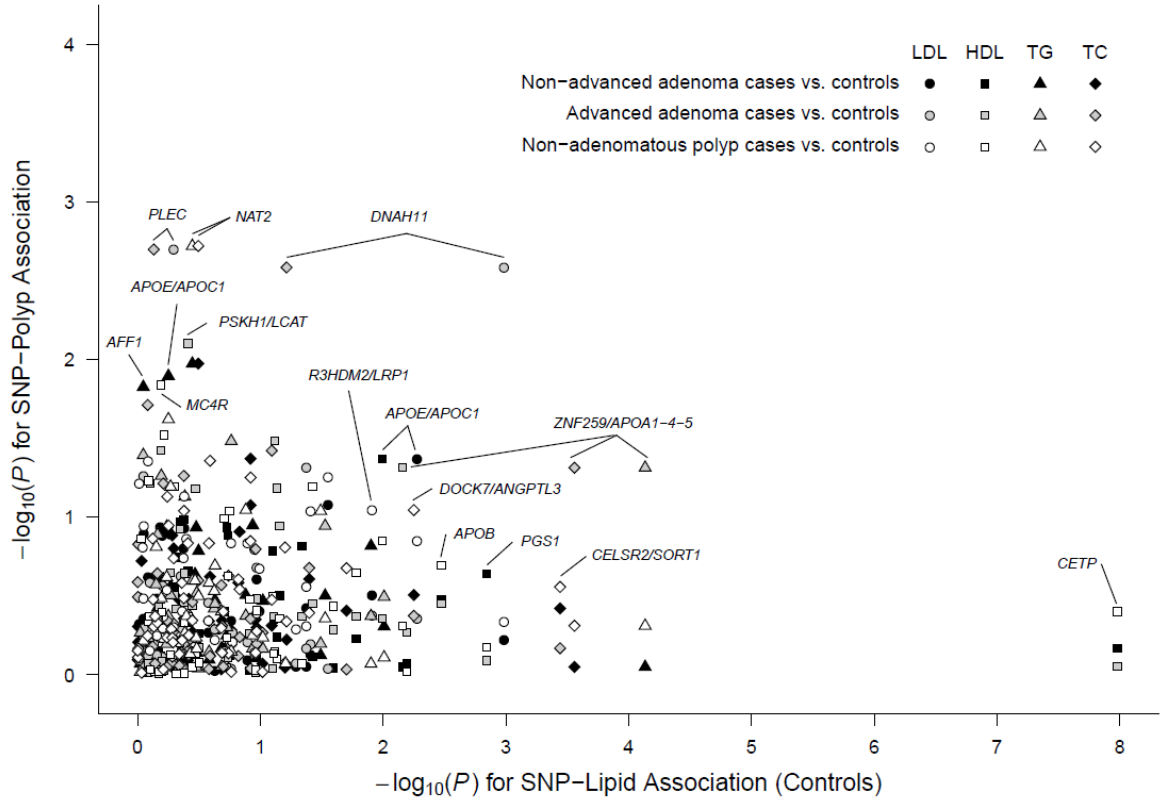


Figure 4.2. Estimated difference in mean value of the blood lipid phenotypes (zenith LDL in A, nadir HDL in B, zenith TG in C, and zenith TC in D) comparing deciles of the allele score in controls to the first decile, Group Health, 1998-2008. Estimated odds ratios of adenomas vs. controls (black circles; plotted with respect to right Y-axis) comparing deciles of the allele score in controls to the first decile, odds ratios of advanced adenomas vs. controls (black triangles), and odds ratios of non-adenomatous polyps vs. controls (white circles). All estimates are adjusted for age, sex, race, and data-collection period; 95% CIs shown. CI, confidence interval; HDL, high-density lipoprotein cholesterol; LDL, low-density lipoprotein cholesterol; TC, total cholesterol; TG triglycerides

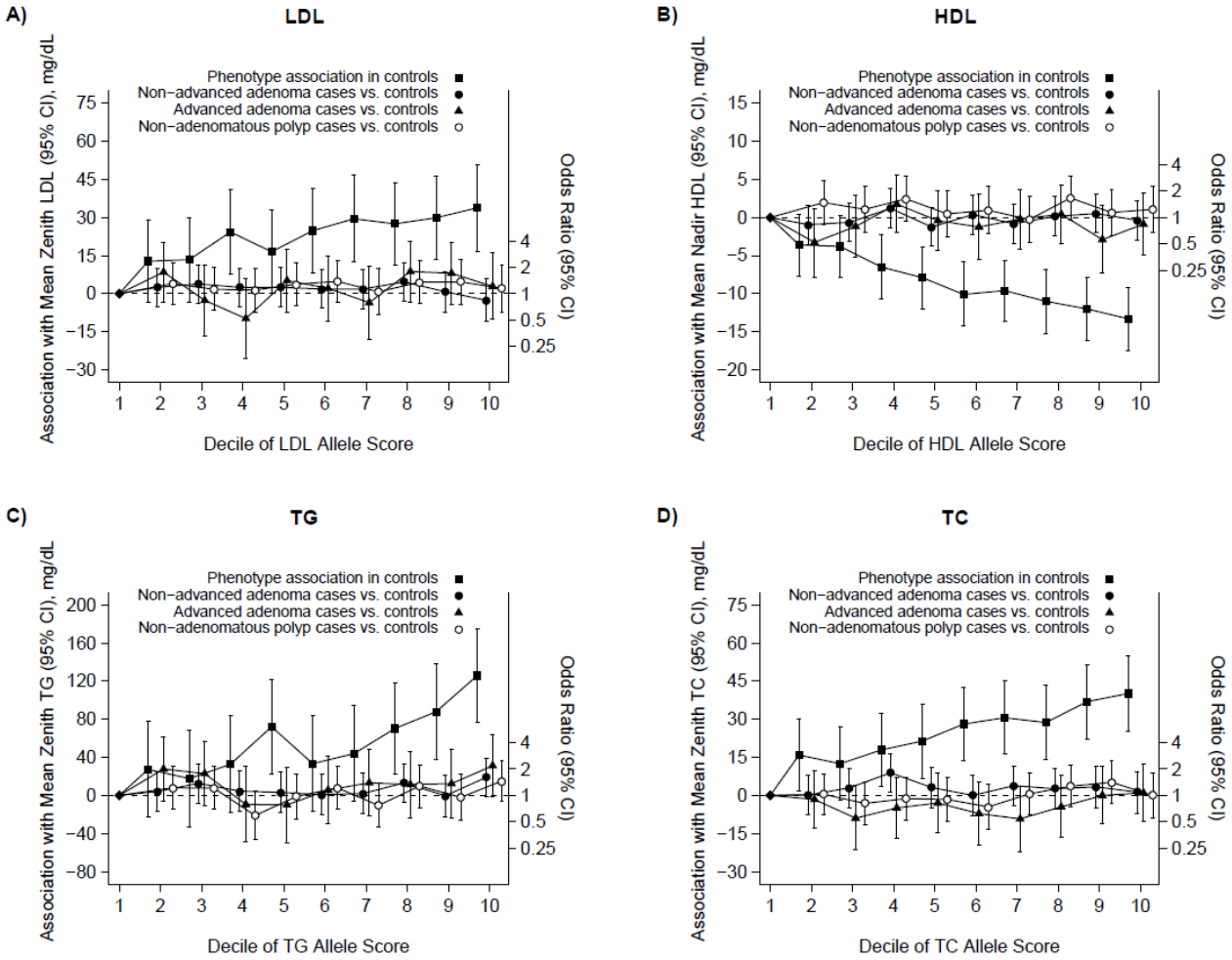
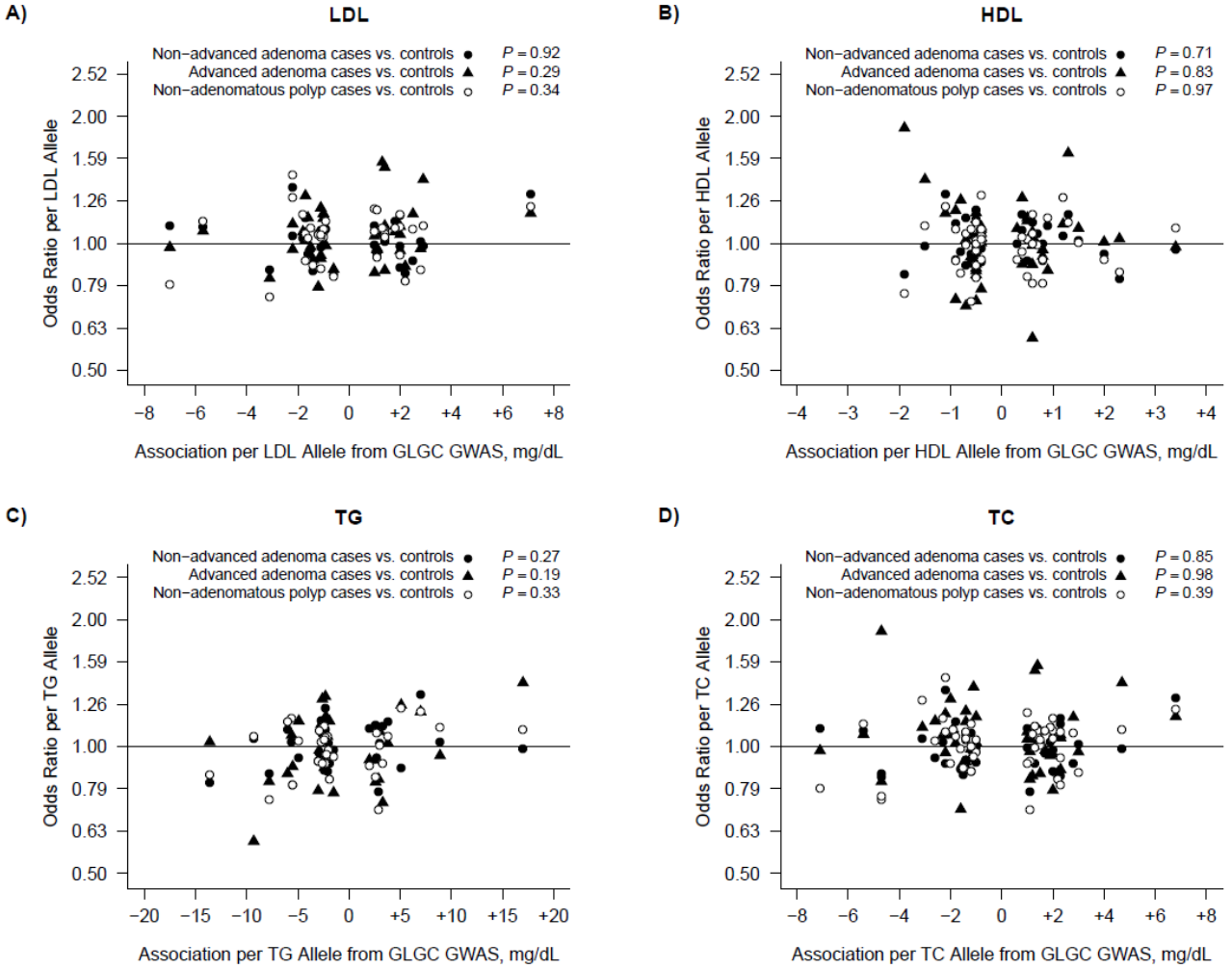


Figure 4.3. Estimated odds ratios of polyps per allele against the association per allele with the lipid component reported by the GLGC GWAS, Group Health, 1998-2007. Estimated odds ratios of adenomas vs. controls (black circles), odds ratios of advanced adenomas vs. controls (black triangles), and odds ratios of non-adenomatous polyps vs. controls (white circles; LDL SNPs in A, HDL SNPs in B, TG SNPs in C, and TC SNPs in D). Odds ratios are adjusted for age, sex, race, and data-collection period. *P*-values are for the outcome-specific inverse-variance weighted slope. GLGC, Global Lipids Genetic Consortium; GWAS, genome-wide association study; HDL, high-density lipoprotein cholesterol; LDL, low-density lipoprotein cholesterol; SNP, single-nucleotide polymorphism; TC, total cholesterol; TG triglycerides



Supplemental Table 4.1. Sensitivity analyses for LDL-polyp associations accounting for use of lipid-controlling drugs.

			Non-advanced adenomas vs. controls		Advanced adenomas vs. controls		Non-adenomatous polyps vs. controls	
Phenotype[a]	Adjustment[b]	Unit change for OR	OR (95% CI)	*P*	OR (95% CI)	*P*	OR (95% CI)	*P*
All lipid-controlling drug users excluded[c]								
Zenith LDL	Minimal	+20 mg/dL	1.15 (1.04, 1.27)	0.006	1.19 (1.04, 1.39)	0.03	1.13 (1.01, 1.27)	0.03
Zenith LDL	Full	+20 mg/dL	1.15 (1.03, 1.29)	0.02	1.26 (1.05, 1.51)	0.01	1.13 (1.00, 1.28)	0.05
Lipid-controlling drug users with pre-colonoscopy zenith LDL<130 mg/dL excluded[d]								
Zenith LDL	Minimal	+20 mg/dL	1.06 (0.99, 1.14)	0.10	1.13 (1.01, 1.27)	0.03	1.06 (0.97, 1.14)	0.19
Zenith LDL	Full	+20 mg/dL	1.06 (0.98, 1.15)	0.17	1.18 (1.04, 1.34)	0.009	1.10 (1.00, 1.20)	0.04
Lipid-controlling drug users with pre-colonoscopy zenith LDL<130 mg/dL imputed[e]								
Zenith LDL	Minimal	+20 mg/dL	1.05 (0.99, 1.13)	0.16	1.12 (1.00, 1.25)	0.05	1.05 (0.97, 1.13)	0.27
Zenith LDL	Full	+20 mg/dL	1.05 (0.97, 1.14)	0.22	1.16 (1.03, 1.31)	0.02	1.09 (1.00, 1.19)	0.05

[a] Polytomous logistic regression model with case-control status as the dependent variable and the highest LDL measurement in at most 20 years prior to colonoscopy as the independent variable.

[b] Minimally-adjusted OR is adjusted for age at colonoscopy, sex, race, and data-collection period. Fully-adjusted OR is adjusted for age at colonoscopy, sex, race, education, BMI, NSAID use, family history of colorectal cancer, estrogen-only use (in women), estrogen-plus-progestin use (in women), cigarette smoking, alcohol consumption, diabetes mellitus, fruit servings per day, vegetable servings per day, recreational and exercise physical activity, prior endoscopy (approximately 2 years before study colonoscopy), and data-collection period.

[c] N=317 controls, N=224 non-advanced adenoma cases, N=60 advanced adenoma cases, N=171 non-adenomatous polyp cases.

[d] N=477 controls, N=362 non-advanced adenoma cases, N=94 advanced adenoma cases, N=252 non-adenomatous polyp cases.

[e] N=513 controls, N=390 non-advanced adenoma cases, N=103 advanced adenoma cases, N=269 non-adenomatous polyp cases. Single-imputed value = max(100 mg/dL, observed zenith LDL + 30 mg/dL).

Abbreviations: BMI, body mass index; CI, confidence interval; LDL, low-density lipoprotein cholesterol; NSAID, nonsteroidal anti-inflammatory drug; OR, odds ratio.

Supplemental Table 4.2. Summary of 96 of the 102 single-nucleotide polymorphisms (SNPs) identified by the Global Lipids Genetics Consortium genome-wide association study (GLGC GWAS).

			Phenotype achieving genome-wide significance from GLGC GWAS		MAF (%)			
SNP[a]	Nearest	Region	Strongest	Other	Controls (N=754)	Non-advanced adenomas (N=518)	Advanced adenomas (N=139)	Non-adenomatous polyps (N=380)
rs629301 (T>G)	*CELSR2/SORT1*	1p21-p13.1	LDL	TC	21	23	23	23
rs989653 (G>A)[b]	*EVI5*	1p22.1	TC		21	19	20	22
rs2131925 (T>G)	*DOCK7/ANGPTL3*	1p31.3	TG	LDL, TC	33	32	36	33
rs2479409 (A>G)	*PCSK9*	1p32.3	LDL	TC	35	33	37	39
rs4660293 (A>G)	*PABPC4*	1p34.2	HDL		23	22	20	24
rs12027135 (T>A)	*TMEM57/LDLRAP1*	1p36-p35	TC	LDL	48	47	46	43
rs1689800 (A>G)	*ZNF648*	1q25.3	HDL		37	35	33	36
rs2642442 (T>C)	*MARC1*	1q41	TC	LDL	31	32	35	32
rs4846914 (A>G)	*GALNT2*	1q41-q42	HDL	TG	44	42	42	45
rs514230 (A>T)	*IRF2BP2*	1q42.3	TC	LDL	45	46	43	46
rs4299376 (T>G)	*ABCG8*	2p21	LDL	TC	31	31	29	29
rs1260326 (C>T)	*GCKR*	2p23	TG	TC	38	39	37	42
rs4564803 (G>T)[c]	*APOB*	2p24-p23	TG	HDL	25	27	23	27
rs7570971 (C>A)	*RAB3GAP1*	2q21.3	TC		40	42	43	39
rs10195252 (T>C)	*COBLL1*	2q24.3	TG		39	39	38	37
rs12328675 (T>C)	*COBLL1*	2q24.3	HDL		11	12	11	11
rs2972146 (T>G)	*IRS1*	2q36	HDL	TG	35	33	38	32
rs2290159 (G>C)	*RAF1*	3p25	TC		21	20	19	22
rs645040 (T>G)	*MSL2*	3q22.3	TG		21	22	24	20
rs442177 (T>G)	*AFF1*	4q21	TG		37	42	44	39
rs13107325 (C>T)	*SLC39A8*	4q22-q24	HDL		6	6	8	6
rs6450176 (G>A)	*ARL15*	5p15.2	HDL		28	30	31	29
rs9686661 (C>T)	*MAP3K1*	5q11.2	TG		20	22	17	18
rs12916 (T>C)	*HMGCR*	5q13.3-q14	TC	LDL	36	34	41	39
rs6882076 (C>T)	*TIMD4*	5q33.3	TC	LDL, TG	36	38	42	35
rs1800562 (G>A)	*HFE*	6p21.3	LDL	TC	5	6	5	7

rs2247056 (C>T)	*HLA-C*	6p21.3	TG		26	25	22	25
rs3177928 (G>A)	*HLA-DRA*	6p21.3	TC	LDL	13	14	13	13
rs2814944 (G>A)	*C6orf106*	6p21.31	HDL		14	17	17	14
rs2814982 (C>T)	*C6orf106*	6p21.31	TC		11	12	12	11
rs3757354 (C>T)	*MYLIP*	6p23-p22.3	LDL	TC	24	22	23	23
rs495565 (A>G)[d]	*FRK*	6q21-q22.3	TC	LDL	35	37	35	38
rs605066 (T>C)	*CITED2*	6q23.3	HDL		46	47	46	48
rs1564348 (T>C)	*SLC22A1/LPA*	6q25.3-q26	LDL	TC	15	14	14	14
rs1084651 (G>A)	*LPA*	6q26	HDL		19	19	20	18
rs2072183 (G>C)	*NPC1L1*	7p13	TC	LDL	25	25	21	25
rs12670798 (T>C)	*DNAH11*	7p21	TC	LDL	23	24	32	23
rs17145738 (C>T)	*TBL2/MLXIPL*	7q11.23	TG	HDL	12	13	8	13
rs4731702 (C>T)	*KLF14*	7q32.3	HDL		47	50	44	48
rs12678919 (A>G)	*LPL*	8p22	TG	HDL	10	9	10	9
rs1495741 (A>G)	*NAT2*	8p22	TG	TC	28	25	26	23
rs11776767 (G>C)	*PINX1*	8p23	TG		39	41	37	35
rs9987289 (G>A)	*PPP1R3B*	8p23.1	HDL	LDL, TC	7	7	8	9
rs2081687 (C>T)	*CYP7A1*	8q11-q12	TC	LDL	32	32	29	34
rs11136341 (A>G)	*PLEC*	8q24	LDL	TC	35	36	45	37
rs2293889 (G>T)	*TRPS1*	8q24.12	HDL		42	43	36	44
rs2737229 (A>C)	*TRPS1*	8q24.12	TC		30	30	38	29
rs2954029 (A>T)	*TRIB1*	8q24.13	TG	LDL, HDL, TC	46	46	47	49
rs581080 (C>G)	*TTC39B*	9p22.3	HDL	TC	19	17	14	19
rs1883025 (C>T)	*ABCA1*	9q31.1	HDL	TC	27	25	30	28
rs635634 (C>T)[e]	*ABO*	9q34.2	LDL	TC	21	19	19	18
rs10761731 (A>T)	*JMJD1C*	10q21.3	TG		43	40	41	46
rs2068888 (G>A)	*CYP26A1*	10q23-q24	TG		45	45	49	37
rs2255141 (G>A)	*GPAM*	10q25.2	TC	LDL	29	31	28	28
rs3136441 (T>C)	*F2/LRP4*	11p11-p11.2	HDL		16	15	16	13
rs2923084 (A>G)	*AMPD3*	11p15	HDL		20	20	22	19
rs10128711 (C>T)	*SPTY2D1*	11p15.1	TC		32	30	33	30
rs174546 (C>T)	*FADS1*	11q12.2-q13.1	TG	LDL, HDL, TC	33	36	33	35
rs964184 (C>G)	*ZNF259/APOA1-4-5*	11q23-q24	TG	LDL, HDL, TC	14	14	19	15

rs7941030 (T>C)	*UBASH3B*	11q24.1	TC	HDL	39	39	41	36
rs11220462 (G>A)	*ST3GAL4*	11q24.2	LDL	TC	14	14	14	15
rs7134375 (C>A)	*PDE3A*	12p12	HDL		40	43	45	39
rs11613352 (C>T)	*R3HDM2/LRP1*	12q13.3	TG	HDL	22	24	23	22
rs11065987 (A>G)	*BRAP*	12q24	TC	LDL	40	39	43	41
rs7134594 (T>C)	*MMAB/MVK*	12q24	HDL		47	48	49	48
rs1169288 (A>C)	*HNF1A*	12q24.2	TC	LDL	32	32	33	33
rs4765127 (G>T)	*ZNF664*	12q24.31	HDL	TG	33	34	30	34
rs838880 (T>C)	*SCARB1*	12q24.31	HDL		35	36	35	31
rs8017377 (G>A)	*NYNRIN*	14q12	LDL		45	43	43	50
rs2412710 (G>A)	*CAPN3*	15q15.1	TG		2	2	2	2
rs2929282 (A>T)	*FRMD5*	15q15.3	TG		6	5	7	6
rs1532085 (G>A)	*LIPC*	15q21-q23	HDL	TC, TG	40	40	42	39
rs11649653 (C>G)	*CTF1*	16p11.2	TG		42	40	44	44
rs3764261 (C>A)	*CETP*	16q21	HDL	LDL, TG, TC	31	30	30	31
rs16942887 (G>A)	*PSKH1/LCAT*	16q22.1	HDL		11	13	17	12
rs2000999 (G>A)	*HPR*	16q22.1	TC	LDL	18	21	19	18
rs2925979 (C>T)	*CMIP*	16q23	HDL		31	31	28	28
rs11869286 (C>G)	*STARD3*	17q11-q12	HDL		34	37	33	36
rs7206971 (G>A)	*EFCAB13*	17q21.32	LDL	TC	46	48	47	51
rs4148008 (C>G)	*ABCA8*	17q24	HDL		34	34	36	33
rs4129767 (G>A)[f]	*PGS1*	17q25.3	HDL		46	49	45	46
rs10871777 (A>G)[g]	*MC4R*	18q22	HDL		24	23	25	28
rs10401969 (T>C)	*SUGP1/CILP2*	19p13.11	TC	LDL, TG	9	8	8	7
rs6511720 (G>T)	*LDLR*	19p13.2	LDL	TC	11	11	10	9
rs737337 (T>C)	*DOCK6*	19p13.2	HDL		10	9	11	7
rs1044250 (C>T)[h]	*ANGPTL4*	19p13.3	HDL		31	31	29	29
rs439401 (C>T)	*APOE/APOC1*	19q13.2	TG		40	36	37	36
rs4420638 (A>G)	*APOE/APOC1*	19q13.2	LDL	HDL, TC	16	19	18	19
rs492602 (A>G)	*FUT2*	19q13.3	TC		47	44	47	51
rs386000 (G>C)	*LILRA3*	19q13.4	HDL		24	24	24	22
rs2902940 (A>G)	*MAFB*	20q11.2-q13.1	TC	LDL	28	29	31	30
rs6029526 (T>A)	*TOP1*	20q12-q13.1	LDL	TC	50	51	46	50

rs1800961 (C>T)	*HNF4A*	20q13.12	HDL	TC	3	3	6	3
rs6065906 (T>C)	*PLTP/PCIF1*	20q13.12	HDL	TG	16	18	13	16
rs181362 (C>T)	*UBE2L3*	22q11.21	HDL		23	24	19	22
rs5756931 (T>C)	*PLA2G6*	22q13.1	TG		37	37	32	37

[a] Of 102 SNPs (major allele > minor allele) from GLGC GWAS, 6 were not genotyped: rs1367117 (*APOB*), rs13238203 (*TYW1B*), rs4759375 (*SBNO1*), rs2652834 (*LACTB*), rs7241918 (*LIPG*), rs2277862 (*ERGIC3*).

[b] rs989653 was used as a proxy for rs7515577.

[c] rs4564803 was used as a proxy for rs1042034.

[d] rs495565 was used as a proxy for rs9488822.

[e] rs635634 replaced rs9411489 according to a change made to the NCBI reference sequence.

[f] A allele of rs4129767 is minor in our sample, but G allele is minor from HapMap and GLGC GWAS. We report associations per copy of G allele to be consistent with GLGC GWAS.

[g] rs10871777 was used as a proxy for rs12967135.

[h] rs1044250 was used as a proxy for rs7255436.

Abbreviations: GLGC, Global Lipids Genetics Consortium; GWAS, genome-wide association study; HDL, high-density lipoprotein cholesterol; LDL, low-density lipoprotein cholesterol; MAF, minor-allele frequency; NCBI, National Center for Biotechnology Information; SNP, single-nucleotide polymorphism; TC, total cholesterol; TG, triglycerides.

GLGC GWAS results are from Teslovich TM, Musunuru K, Smith AV, et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* 2010;466(7307):707-713.

Supplemental Table 4.3. Genotype-lipid (among controls) and genotype-polyp associations for the single-SNPs associated with LDL. SNPs with $P \leq 0.01$ for genotype-lipid association among controls shown in bold.

				Zenith LDL among controls from Group Health				Non-advanced adenomas vs. controls		Advanced adenomas vs. controls		Non-adenomatous polyps vs. controls	
SNP[a]	MAF (%)[b]	Nearest gene(s)	LDL association from GLGC GWAS[c]	Association[d] (95% CI)	Direction/ Coverage[e]	*F*	*P*	OR[f] (95% CI)	*P*	OR[f] (95% CI)	*P*	OR[f] (95% CI)	*P*
rs6511720 (G>T)	11	*LDLR*	-7.0	-6 (-14, +3)	Y/Y	1.9	0.17	1.10 (0.86, 1.42)	0.46	0.98 (0.64, 1.49)	0.92	0.80 (0.59, 1.08)	0.15
rs629301 (T>G)	21	*CELSR2/ SORT1*	-5.7	-7 (-14, 0)	Y/Y	4.2	0.04	1.09 (0.90, 1.33)	0.38	1.07 (0.78, 1.46)	0.68	1.13 (0.91, 1.40)	0.28
rs10401969 (T>C)	9	*SUGP1/ CILP2*	-3.1	-12 (-21, -3)	Y/Y	6.3	0.01	0.86 (0.65, 1.15)	0.31	0.83 (0.52, 1.32)	0.42	0.75 (0.53, 1.05)	0.09
rs1800562 (G>A)	5	*HFE*	-2.2	-14 (-26, -2)	Y/Y	4.8	0.03	1.36 (0.96, 1.92)	0.08	0.97 (0.53, 1.77)	0.91	1.45 (0.99, 2.13)	0.06
rs9987289 (G>A)	7	*PPP1R3B*	-2.2	+5 (-6, +15)	N/Y	0.7	0.40	1.04 (0.77, 1.41)	0.78	1.11 (0.69, 1.78)	0.66	1.28 (0.92, 1.78)	0.14
rs2954029 (A>T)	46	*TRIB1*	-1.8	-6 (-11, 0)	Y/Y	4.3	0.04	1.03 (0.88, 1.20)	0.75	1.06 (0.82, 1.37)	0.63	1.17 (0.98, 1.40)	0.09
rs174546 (C>T)	33	*FADS1*	-1.7	+1 (-4, +7)	N/Y	0.2	0.66	1.15 (0.97, 1.36)	0.12	1.02 (0.78, 1.34)	0.89	1.06 (0.87, 1.28)	0.58
rs6882076 (C>T)	36	*TIMD4*	-1.7	0 (-6, +5)	Y/Y	0.0	0.90	1.07 (0.90, 1.26)	0.44	1.30 (1.00, 1.69)	0.05	0.91 (0.75, 1.10)	0.33
rs2131925 (T>G)	33	*DOCK7/ ANGPTL3*	-1.6	0 (-6, +6)	N/Y	0.0	0.98	0.94 (0.79, 1.12)	0.49	1.15 (0.87, 1.51)	0.32	1.03 (0.85, 1.25)	0.78
rs3764261 (C>A)	31	*CETP*	-1.5	+2 (-3, +8)	N/Y	0.6	0.43	0.96 (0.81, 1.15)	0.68	0.98 (0.74, 1.30)	0.89	1.09 (0.90, 1.33)	0.40
rs3757354 (C>T)	24	*MYLIP*	-1.4	-2 (-8, +5)	Y/Y	0.3	0.58	0.86 (0.70, 1.04)	0.12	0.92 (0.67, 1.25)	0.58	0.89 (0.71, 1.11)	0.29
rs2072183 (G>C)	25	*NPC1L1*	-1.2	0 (-6, +6)	N/Y	0.0	0.97	1.05 (0.87, 1.27)	0.62	0.79 (0.57, 1.09)	0.15	1.05 (0.85, 1.29)	0.68
rs12027135 (T>A)	48	*TMEM57/ LDLRAP1*	-1.1	+4 (-1, +9)	N/Y	2.3	0.13	0.98 (0.83, 1.15)	0.81	0.94 (0.72, 1.21)	0.61	0.87 (0.73, 1.05)	0.15
rs2642442 (T>C)	31	*MARC1*	-1.1	-5 (-11, +1)	Y/Y	2.5	0.11	1.07 (0.90, 1.27)	0.44	1.21 (0.93, 1.59)	0.16	1.04 (0.85, 1.26)	0.72
rs514230 (A>T)	45	*IRF2BP2*	-1.1	-4 (-9, +2)	Y/Y	1.9	0.17	1.05 (0.89, 1.23)	0.58	0.92 (0.71, 1.20)	0.54	1.05 (0.87, 1.25)	0.62
rs11065987 (A>G)	40	*BRAP*	-1.0	-2 (-7, +3)	Y/Y	0.6	0.45	0.99 (0.84, 1.17)	0.91	1.18 (0.90, 1.54)	0.23	1.03 (0.86, 1.25)	0.73
rs2902940 (A>G)	28	*MAFB*	-1.0	-3 (-9, +3)	Y/Y	1.3	0.27	1.06 (0.89, 1.27)	0.53	1.15 (0.86, 1.52)	0.35	1.08 (0.88, 1.32)	0.45
rs495565 (A>G)[g]	35	*FRK*	-0.9	-5 (-10, +1)	Y/Y	2.7	0.10	1.08 (0.91, 1.27)	0.39	0.99 (0.75, 1.29)	0.92	1.13 (0.93, 1.36)	0.21
rs1564348 (T>C)	15	*SLC22A1/ LPA*	-0.6	-3 (-10, +4)	Y/Y	0.6	0.42	0.84 (0.66, 1.07)	0.16	0.87 (0.59, 1.28)	0.47	0.83 (0.64, 1.09)	0.18
rs2081687 (C>T)	32	*CYP7A1*	+1.0	+1 (-4, +7)	Y/Y	0.2	0.63	0.99 (0.83, 1.18)	0.92	0.85 (0.64, 1.13)	0.27	1.07 (0.88, 1.30)	0.50
rs7206971 (G>A)	46	*EFCAB13*	+1.0	-1 (-6, +5)	N/Y	0.1	0.83	1.10 (0.94, 1.29)	0.24	1.04 (0.81, 1.35)	0.76	1.20 (1.01, 1.44)	0.04
rs2255141 (G>A)	29	*GPAM*	+1.1	+2 (-4, +7)	Y/Y	0.3	0.62	1.07 (0.90, 1.27)	0.43	0.97 (0.74, 1.29)	0.85	0.92 (0.76, 1.12)	0.43
rs8017377 (G>A)	45	*NYNRIN*	+1.1	0 (-5, +5)	N/Y	0.0	0.97	0.94 (0.79, 1.11)	0.47	0.97 (0.74, 1.27)	0.81	1.20 (0.99, 1.45)	0.06
rs12670798 (T>C)	**23**	***DNAH11***	**+1.3**	**+11 (+4, +17)**	**Y/N**	**10.9**	**0.001**	**1.05 (0.87, 1.28)**	**0.60**	**1.56 (1.17, 2.08)**	**0.003**	**1.09 (0.87, 1.35)**	**0.46**

rs11136341 (A>G)	35	*PLEC*	+1.4	-2 (-7, +4)	N/Y	0.4	0.51	1.06 (0.90, 1.26)	0.49	1.52 (1.17, 1.97)	0.002	1.07 (0.88, 1.29)	0.50
rs1169288 (A>C)	32	*HNF1A*	+1.4	+4 (-2, +9)	Y/Y	1.4	0.24	1.01 (0.84, 1.21)	0.95	1.10 (0.83, 1.46)	0.51	1.06 (0.86, 1.29)	0.60
rs6029526 (T>A)	50	*TOP1*	+1.4	+1 (-5, +6)	Y/Y	0.1	0.79	1.02 (0.87, 1.20)	0.80	0.86 (0.67, 1.11)	0.26	1.04 (0.87, 1.24)	0.70
rs3177928 (G>A)	13	*HLA-DRA*	+1.8	-2 (-9, +6)	N/Y	0.2	0.70	1.13 (0.89, 1.43)	0.33	1.07 (0.72, 1.57)	0.75	1.08 (0.83, 1.42)	0.56
rs11220462 (G>A)	14	*ST3GAL4*	+2.0	+8 (0, +15)	Y/Y	3.8	0.05	0.98 (0.78, 1.25)	0.88	0.96 (0.65, 1.41)	0.84	1.09 (0.84, 1.41)	0.51
rs2000999 (G>A)	18	*HPR*	+2.0	+2 (-5, +8)	Y/Y	0.2	0.63	1.16 (0.96, 1.42)	0.13	1.05 (0.76, 1.44)	0.77	0.94 (0.75, 1.18)	0.58
rs2479409 (A>G)	35	*PCSK9*	+2.0	0 (-6, +5)	N/Y	0.0	0.89	0.87 (0.74, 1.04)	0.13	1.10 (0.84, 1.43)	0.50	1.17 (0.96, 1.41)	0.11
rs635634 (C>T)[h]	21	*ABO*	+2.2	-3 (-9, +4)	N/Y	0.7	0.42	0.85 (0.69, 1.04)	0.12	0.88 (0.64, 1.22)	0.45	0.81 (0.65, 1.02)	0.07
rs12916 (T>C)	36	*HMGCR*	+2.5	+4 (-1, +9)	Y/Y	2.6	0.11	0.91 (0.77, 1.07)	0.25	1.17 (0.91, 1.51)	0.21	1.08 (0.90, 1.29)	0.40
rs4299376 (T>G)	31	*ABCG8*	+2.8	0 (-5, +6)	Y/Y	0.0	0.91	1.01 (0.85, 1.21)	0.90	0.97 (0.73, 1.29)	0.84	0.87 (0.71, 1.06)	0.16
rs964184 (C>G)	14	*ZNF259/ APOA1-4-5*	+2.9	+8 (0, +15)	Y/Y	4.2	0.04	0.98 (0.78, 1.25)	0.89	1.42 (1.00, 2.00)	0.05	1.10 (0.85, 1.42)	0.49
rs4420638 (A>G)	**16**	***APOE/ APOC1***	**+7.1**	**+12 (+4, +21)**	**Y/Y**	**7.9**	**0.005**	**1.30 (1.01, 1.69)**	**0.04**	**1.18 (0.78, 1.79)**	**0.44**	**1.22 (0.94, 1.60)**	**0.14**

[a] SNPs (major allele > minor allele) associated with LDL at $P < 5\times10^{-8}$ from GLGC GWAS.

[b] MAF among controls.

[c] mg/dL increase (or decrease) in mean LDL per minor allele reported by GLGC GWAS.

[d] mg/dL increase (or decrease) in zenith LDL (mg/dL) per minor allele among controls from linear regression model adjusted for age at colonoscopy, sex, race, and data-collection period.

[e] Y if associations from GLGC and controls have same sign before rounding; N otherwise / Y if association from GLGC is contained within the 95% CI from controls before rounding; N otherwise.

[f] OR from polytomous logistic regression model adjusted for age at colonoscopy, sex, race, and data-collection period.

[g] rs495565 was used as a proxy for rs9488822.

[h] rs635634 replaced rs9411489 according to a change made to the NCBI reference sequence.

Abbreviations: CI, confidence interval; GLGC, Global Lipids Genetics Consortium; GWAS, genome-wide association study; LDL, low-density lipoprotein cholesterol; MAF, minor-allele frequency; N, no; NCBI, National Center for Biotechnology Information; OR, odds ratio; SNP, single-nucleotide polymorphism; Y, yes.

GLGC GWAS results are from Teslovich TM, Musunuru K, Smith AV, et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* 2010;466(7307):707-713.

Supplemental Table 4.4. Genotype-lipid (among controls) and genotype-polyp associations for the single-SNPs associated with HDL. SNPs with $P \leq 0.01$ for genotype-lipid association among controls shown in bold.

				Nadir HDL among controls from Group Health				Non-advanced adenomas vs. controls		Advanced adenomas vs. controls		Non-adenomatous polyps vs. controls	
SNP[a]	MAF (%)[b]	Nearest gene(s)	HDL association from GLGC GWAS[c]	Association[d] (95% CI)	Direction/ Coverage[e]	*F*	*P*	OR[f] (95% CI)	*P*	OR[f] (95% CI)	*P*	OR[f] (95% CI)	*P*
rs1800961 (C>T)	3	*HNF4A*	-1.9	+1 (-3, +5)	N/Y	0.2	0.65	0.84 (0.52, 1.37)	0.49	1.88 (1.04, 3.40)	0.04	0.76 (0.44, 1.33)	0.33
rs964184 (C>G)	**14**	***ZNF259/ APOA1-4-5***	**-1.5**	**-3 (-5, -1)**	**Y/Y**	**7.3**	**0.007**	**0.98 (0.78, 1.25)**	**0.89**	**1.42 (1.00, 2.00)**	**0.05**	**1.10 (0.85, 1.42)**	**0.49**
rs4420638 (A>G)	**16**	***APOE/APOC1***	**-1.1**	**-3 (-5, -1)**	**Y/Y**	**6.7**	**0.01**	**1.30 (1.01, 1.69)**	**0.04**	**1.18 (0.78, 1.79)**	**0.44**	**1.22 (0.94, 1.60)**	**0.14**
rs1883025 (C>T)	27	*ABCA1*	-0.9	+1 (-1, +2)	N/Y	0.7	0.42	0.91 (0.76, 1.10)	0.34	1.20 (0.90, 1.60)	0.23	1.08 (0.88, 1.34)	0.47
rs6065906 (T>C)	16	*PLTP/PCIF1*	-0.9	-2 (-3, 0)	Y/Y	3.3	0.07	1.11 (0.90, 1.38)	0.31	0.74 (0.50, 1.08)	0.11	0.91 (0.71, 1.16)	0.44
rs13107325 (C>T)	6	*SLC39A8*	-0.8	-2 (-5, +1)	Y/Y	2.6	0.11	0.95 (0.69, 1.32)	0.77	1.27 (0.79, 2.04)	0.33	0.85 (0.59, 1.23)	0.39
rs174546 (C>T)	33	*FADS1*	-0.7	-1 (-2, +1)	Y/Y	1.8	0.19	1.15 (0.97, 1.36)	0.12	1.02 (0.78, 1.34)	0.89	1.06 (0.87, 1.28)	0.58
rs581080 (C>G)	19	*TTC39B*	-0.7	0 (-1, +2)	N/Y	0.1	0.79	0.88 (0.72, 1.09)	0.24	0.71 (0.50, 1.02)	0.06	0.99 (0.79, 1.24)	0.92
rs4846914 (A>G)	44	*GALNT2*	-0.6	0 (-2, +1)	Y/Y	0.1	0.79	0.94 (0.80, 1.11)	0.47	0.93 (0.72, 1.21)	0.58	1.08 (0.90, 1.30)	0.42
rs737337 (T>C)	10	*DOCK6*	-0.6	0 (-2, +3)	N/Y	0.1	0.81	0.90 (0.68, 1.18)	0.43	1.04 (0.69, 1.58)	0.85	0.73 (0.52, 1.01)	0.06
rs1044250 (C>T)[g]	31	*ANGPTL4*	-0.5	+2 (0, +3)	N/N	5.8	0.02	0.95 (0.80, 1.13)	0.59	0.89 (0.67, 1.18)	0.42	0.89 (0.73, 1.08)	0.22
rs11869286 (C>G)	34	*STARD3*	-0.5	-1 (-2, +1)	Y/Y	1.8	0.18	1.14 (0.96, 1.35)	0.13	0.96 (0.73, 1.27)	0.78	1.12 (0.93, 1.36)	0.24
rs1689800 (A>G)	37	*ZNF648*	-0.5	0 (-1, +2)	N/Y	0.4	0.54	0.94 (0.79, 1.12)	0.48	0.84 (0.64, 1.11)	0.22	0.97 (0.80, 1.17)	0.75
rs181362 (C>T)	23	*UBE2L3*	-0.5	-1 (-3, +1)	Y/Y	0.9	0.34	1.03 (0.85, 1.25)	0.75	0.73 (0.52, 1.02)	0.07	0.96 (0.77, 1.20)	0.73
rs2814944 (G>A)	14	*C6orf106*	-0.5	-1 (-3, +1)	Y/Y	0.7	0.42	1.20 (0.96, 1.50)	0.10	1.18 (0.83, 1.68)	0.35	1.00 (0.77, 1.29)	0.98
rs2925979 (C>T)	31	*CMIP*	-0.5	-2 (-3, 0)	Y/Y	4.3	0.04	0.97 (0.82, 1.16)	0.77	0.87 (0.66, 1.16)	0.35	0.83 (0.67, 1.01)	0.06
rs4660293 (A>G)	23	*PABPC4*	-0.5	-1 (-2, +1)	Y/Y	0.8	0.36	0.96 (0.79, 1.17)	0.71	0.86 (0.62, 1.19)	0.36	1.02 (0.82, 1.25)	0.89
rs6450176 (G>A)	28	*ARL15*	-0.5	0 (-1, +2)	N/Y	0.2	0.67	1.04 (0.86, 1.24)	0.70	1.14 (0.86, 1.52)	0.37	0.99 (0.81, 1.22)	0.95
rs10871777 (A>G)[h]	24	*MC4R*	-0.4	0 (-2, +1)	Y/Y	0.2	0.65	0.97 (0.80, 1.18)	0.76	1.03 (0.76, 1.39)	0.86	1.30 (1.05, 1.60)	0.01
rs2293889 (G>T)	42	*TRPS1*	-0.4	-1 (-3, 0)	Y/Y	3.2	0.07	1.05 (0.89, 1.23)	0.58	0.78 (0.60, 1.02)	0.07	1.03 (0.85, 1.23)	0.79
rs2923084 (A>G)	20	*AMPD3*	-0.4	0 (-2, +1)	Y/Y	0.1	0.73	1.02 (0.83, 1.25)	0.87	1.08 (0.79, 1.49)	0.64	1.01 (0.80, 1.27)	0.96
rs4129767 (G>A)[i]	**54**	***PGS1***	**-0.4**	**-2 (-4, -1)**	**Y/N**	**10.2**	**0.007**	**0.90 (0.77, 1.07)**	**0.23**	**1.03 (0.80, 1.34)**	**0.81**	**1.04 (0.87, 1.25)**	**0.67**
rs4148008 (C>G)	34	*ABCA8*	-0.4	-2 (-3, 0)	Y/Y	5.0	0.03	1.01 (0.85, 1.20)	0.90	1.10 (0.83, 1.44)	0.52	0.91 (0.75, 1.11)	0.37
rs605066 (T>C)	46	*CITED2*	-0.4	0 (-2, +1)	Y/Y	0.2	0.68	1.03 (0.86, 1.23)	0.75	1.00 (0.76, 1.32)	0.98	1.08 (0.88, 1.32)	0.45
rs7134594 (T>C)	47	*MMAB/MVK*	-0.4	-1 (-2, +1)	Y/Y	1.0	0.32	1.05 (0.90, 1.23)	0.55	1.06 (0.82, 1.36)	0.67	1.02 (0.85, 1.22)	0.84
rs7941030 (T>C)	39	*UBASH3B*	+0.3	+1 (0, +2)	Y/Y	2.4	0.12	0.99 (0.84, 1.17)	0.94	1.08 (0.83, 1.41)	0.55	0.91 (0.76, 1.10)	0.33
rs4765127 (G>T)	33	*ZNF664*	+0.4	+1 (0, +3)	Y/Y	3.1	0.08	1.07 (0.90, 1.27)	0.44	0.89 (0.67, 1.18)	0.43	1.03 (0.85, 1.26)	0.74
rs7134375 (C>A)	40	*PDE3A*	+0.4	-1 (-2, +1)	N/Y	0.6	0.45	1.17 (0.97, 1.42)	0.11	1.28 (0.94, 1.76)	0.12	0.95 (0.78, 1.17)	0.63
rs11613352 (C>T)	22	*R3HDM2/ LRP1*	+0.5	+2 (0, +3)	Y/Y	4.0	0.05	1.15 (0.95, 1.39)	0.15	1.13 (0.84, 1.53)	0.42	1.02 (0.82, 1.27)	0.85
rs2972146 (T>G)	35	*IRS1*	+0.5	+1 (-1, +2)	Y/Y	0.5	0.50	0.91 (0.77, 1.08)	0.27	1.15 (0.88, 1.50)	0.31	0.83 (0.69, 1.01)	0.06

rs17145738 (C>T)	12	*TBL2/ MLXIPL*	+0.6	+2 (0, +4)	Y/Y	3.1	0.08	1.05 (0.82, 1.34)	0.72	0.60 (0.37, 0.96)	0.03	1.05 (0.80, 1.38)	0.71
rs2954029 (A>T)	46	*TRIB1*	+0.6	+1 (0, +2)	Y/Y	1.8	0.18	1.03 (0.88, 1.20)	0.75	1.06 (0.82, 1.37)	0.63	1.17 (0.98, 1.40)	0.09
rs4731702 (C>T)	47	*KLF14*	+0.6	+1 (-1, +2)	Y/Y	0.5	0.48	1.12 (0.95, 1.31)	0.18	0.89 (0.69, 1.16)	0.38	1.00 (0.83, 1.20)	0.98
rs838880 (T>C)	35	*SCARB1*	+0.6	0 (-1, +2)	Y/Y	0.3	0.61	1.02 (0.86, 1.22)	0.82	0.99 (0.75, 1.31)	0.94	0.80 (0.66, 0.98)	0.03
rs12328675 (T>C)	11	*COBLL1*	+0.7	+2 (0, +4)	Y/Y	2.6	0.11	1.05 (0.82, 1.35)	0.71	0.99 (0.66, 1.49)	0.97	0.99 (0.74, 1.33)	0.93
rs3136441 (T>C)	16	*F2/LRP4*	+0.8	+1 (-1, +3)	Y/Y	1.7	0.20	0.91 (0.72, 1.14)	0.40	0.93 (0.65, 1.34)	0.71	0.80 (0.61, 1.05)	0.10
rs386000 (G>C)	24	*LILRA3*	+0.8	0 (-2, +1)	N/Y	0.2	0.63	0.99 (0.82, 1.20)	0.95	0.97 (0.71, 1.31)	0.83	0.91 (0.74, 1.13)	0.40
rs4564803 (G>T)[j]	**25**	***APOB***	**+0.9**	**+2 (+1, +4)**	**Y/Y**	**8.7**	**0.003**	**1.10 (0.91, 1.32)**	**0.34**	**0.86 (0.63, 1.18)**	**0.35**	**1.15 (0.93, 1.41)**	**0.20**
rs9987289 (G>A)	7	*PPP1R3B*	+1.2	0 (-3, +3)	N/Y	0.0	0.93	1.04 (0.77, 1.41)	0.78	1.11 (0.69, 1.78)	0.66	1.28 (0.92, 1.78)	0.14
rs16942887 (G>A)	11	*PSKH1/ LCAT*	+1.3	+1 (-1, +3)	Y/Y	0.8	0.39	1.17 (0.91, 1.51)	0.22	1.64 (1.14, 2.35)	0.008	1.12 (0.84, 1.50)	0.43
rs1532085 (G>A)	**40**	***LIPC***	**+1.5**	**+2 (+1, +3)**	**Y/Y**	**7.5**	**0.006**	**1.02 (0.86, 1.20)**	**0.85**	**1.09 (0.83, 1.42)**	**0.54**	**1.01 (0.83, 1.21)**	**0.96**
rs1084651 (G>A)	19	*LPA*	+2.0	0 (-2, +2)	Y/Y	0.1	0.82	0.94 (0.76, 1.16)	0.58	1.01 (0.72, 1.41)	0.97	0.91 (0.72, 1.16)	0.46
rs12678919 (A>G)	10	*LPL*	+2.3	+2 (0, +4)	Y/Y	3.1	0.08	0.82 (0.62, 1.08)	0.16	1.03 (0.67, 1.56)	0.91	0.86 (0.63, 1.16)	0.32
rs3764261 (C>A)	**31**	***CETP***	**+3.4**	**+4 (+3, +6)**	**Y/Y**	**33.5**	**1×10^{-8}**	**0.96 (0.81, 1.15)**	**0.68**	**0.98 (0.74, 1.30)**	**0.89**	**1.09 (0.90, 1.33)**	**0.40**

[a] SNPs (major allele > minor allele) associated with HDL at $P< 5\times10^{-8}$ from GLGC GWAS.

[b] MAF among controls.

[c] mg/dL increase (or decrease) in mean HDL per minor allele reported by GLGC GWAS.

[d] mg/dL increase (or decrease) in nadir HDL (mg/dL) per minor allele among controls from linear regression model adjusted for age at colonoscopy, sex, race, and data-collection period.

[e] Y if associations from GLGC and controls have same sign before rounding; N otherwise / Y if association from GLGC is contained within the 95% CI from controls before rounding; N otherwise.

[f] OR from polytomous logistic regression model adjusted for age at colonoscopy, sex, race, and data-collection period.

[g] rs1044250 was used as a proxy for rs7255436.

[h] rs10871777 was used as a proxy for rs12967135.

[i] A allele of rs4129767 is minor here, but G allele is minor from GLGC GWAS. We report associations per copy of G allele to be consistent with GLGC GWAS.

[j] rs4564803 was used as a proxy for rs1042034.

Abbreviations: CI, confidence interval; GLGC, Global Lipids Consortium; GWAS, genome-wide association study; HDL, high-density lipoprotein cholesterol; MAF, minor-allele frequency; N, no; OR, odds ratio; SNP, single-nucleotide polymorphism; Y, yes.

GLGC GWAS results are from Teslovich TM, Musunuru K, Smith AV, et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* 2010;466(7307):707-713.

Supplemental Table 4.5. Genotype-lipid (among controls) and genotype-polyp associations for the single-SNPs associated with TG. SNPs with $P \leq 0.01$ for genotype-lipid association among controls shown in bold.

				Zenith TG among controls from Group Health				Non-advanced adenomas vs. controls		Advanced adenomas vs. controls		Non-adenomatous polyps vs. controls	
SNP[a]	MAF (%)[b]	Nearest gene(s)	TG association from GLGC GWAS[c]	Association[d] (95% CI)	Direction/ Coverage[e]	*F*	*P*	OR[f] (95% CI)	*P*	OR[f] (95% CI)	*P*	OR[f] (95% CI)	*P*
rs12678919 (A>G)	10	*LPL*	-13.6	-13 (-40, +13)	Y/Y	1.0	0.32	0.82 (0.62, 1.08)	0.16	1.03 (0.67, 1.56)	0.91	0.86 (0.63, 1.16)	0.32
rs17145738 (C>T)	12	*TBL2/MLXIPL*	-9.3	-18 (-44, +8)	Y/Y	1.9	0.17	1.05 (0.82, 1.34)	0.72	0.60 (0.37, 0.96)	0.03	1.05 (0.80, 1.38)	0.71
rs10401969 (T>C)	9	*SUGP1/CILP2*	-7.8	-21 (-48, +6)	Y/Y	2.3	0.13	0.86 (0.65, 1.15)	0.31	0.83 (0.52, 1.32)	0.42	0.75 (0.53, 1.05)	0.09
rs4564803 (G>T)[g]	25	*APOB*	-6.0	-11 (-30, +7)	Y/Y	1.4	0.23	1.10 (0.91, 1.32)	0.34	0.86 (0.63, 1.18)	0.35	1.15 (0.93, 1.41)	0.20
rs2954029 (A>T)	46	*TRIB1*	-5.6	-17 (-33, -2)	Y/Y	4.6	0.03	1.03 (0.88, 1.20)	0.75	1.06 (0.82, 1.37)	0.63	1.17 (0.98, 1.40)	0.09
rs439401 (C>T)	40	*APOE/APOC1*	-5.5	+5 (-12, +21)	N/Y	0.3	0.56	0.81 (0.68, 0.96)	0.01	0.90 (0.69, 1.17)	0.42	0.81 (0.67, 0.97)	0.02
rs2131925 (T>G)	**33**	***DOCK7/ ANGPTL3***	**-4.9**	**-23 (-40, -6)**	**Y/Y**	**6.7**	**0.01**	**0.94 (0.79, 1.12)**	**0.49**	**1.15 (0.87, 1.51)**	**0.32**	**1.03 (0.85, 1.25)**	**0.78**
rs2247056 (C>T)	26	*HLA-C*	-3.0	-4 (-22, +14)	Y/Y	0.2	0.69	0.92 (0.76, 1.11)	0.38	0.79 (0.58, 1.07)	0.13	0.92 (0.75, 1.13)	0.44
rs3764261 (C>A)	31	*CETP*	-2.9	-11 (-28, +6)	Y/Y	1.6	0.21	0.96 (0.81, 1.15)	0.68	0.98 (0.74, 1.30)	0.89	1.09 (0.90, 1.33)	0.40
rs11613352 (C>T)	**22**	***R3HDM2/ LRP1***	**-2.7**	**-25 (-44, -5)**	**Y/N**	**6.3**	**0.01**	**1.15 (0.95, 1.39)**	**0.15**	**1.13 (0.84, 1.53)**	**0.42**	**1.02 (0.82, 1.27)**	**0.85**
rs6882076 (C>T)	36	*TIMD4*	-2.6	+4 (-12, +20)	N/Y	0.2	0.64	1.07 (0.90, 1.26)	0.44	1.30 (1.00, 1.69)	0.05	0.91 (0.75, 1.10)	0.33
rs10761731 (A>T)	43	*JMJD1C*	-2.4	-5 (-20, +11)	Y/Y	0.3	0.58	0.88 (0.74, 1.03)	0.12	0.94 (0.72, 1.22)	0.64	1.12 (0.93, 1.34)	0.24
rs4765127 (G>T)	33	*ZNF664*	-2.4	-8 (-25, +9)	Y/Y	0.8	0.37	1.07 (0.90, 1.27)	0.44	0.89 (0.67, 1.18)	0.43	1.03 (0.85, 1.26)	0.74
rs2068888 (G>A)	45	*CYP26A1*	-2.3	-21 (-53, +10)	Y/Y	1.8	0.19	1.09 (0.81, 1.48)	0.56	1.17 (0.74, 1.87)	0.50	0.91 (0.54, 1.55)	0.73
rs442177 (T>G)	37	*AFF1*	-2.3	+1 (-16, +18)	N/Y	0.0	0.90	1.23 (1.04, 1.45)	0.01	1.32 (1.01, 1.71)	0.04	1.05 (0.87, 1.27)	0.61
rs645040 (T>G)	21	*MSL2*	-2.2	+10 (-10, +29)	N/Y	1.0	0.33	1.04 (0.85, 1.26)	0.72	1.17 (0.87, 1.59)	0.30	0.95 (0.76, 1.20)	0.68
rs11649653 (C>G)	42	*CTF1*	-2.1	+13 (-3, +29)	N/Y	2.5	0.12	0.87 (0.74, 1.03)	0.11	1.06 (0.81, 1.37)	0.68	1.06 (0.88, 1.27)	0.56
rs10195252 (T>C)	39	*COBLL1*	-2.0	-16 (-32, +1)	Y/Y	3.5	0.06	1.02 (0.86, 1.20)	0.86	0.98 (0.75, 1.28)	0.86	0.98 (0.81, 1.19)	0.85
rs2972146 (T>G)	35	*IRS1*	-1.9	-5 (-21, +11)	Y/Y	0.4	0.54	0.91 (0.77, 1.08)	0.27	1.15 (0.88, 1.50)	0.31	0.83 (0.69, 1.01)	0.06
rs5756931 (T>C)	37	*PLA2G6*	-1.5	-7 (-23, +9)	Y/Y	0.7	0.41	0.98 (0.83, 1.16)	0.83	0.78 (0.59, 1.03)	0.07	0.94 (0.78, 1.14)	0.54
rs11776767 (G>C)	39	*PINX1*	+2.0	+8 (-8, +23)	Y/Y	0.9	0.34	1.11 (0.94, 1.30)	0.23	0.93 (0.71, 1.21)	0.59	0.90 (0.75, 1.08)	0.25
rs9686661 (C>T)	20	*MAP3K1*	+2.6	+4 (-16, +23)	Y/Y	0.1	0.70	1.12 (0.92, 1.37)	0.25	0.82 (0.59, 1.16)	0.27	0.84 (0.67, 1.07)	0.15
rs4846914 (A>G)	44	*GALNT2*	+2.7	+7 (-9, +23)	Y/Y	0.7	0.40	0.94 (0.80, 1.11)	0.47	0.93 (0.72, 1.21)	0.58	1.08 (0.90, 1.30)	0.42
rs1495741 (A>G)	28	*NAT2*	+2.9	+8 (-9, +26)	Y/Y	0.9	0.36	0.78 (0.65, 0.94)	0.01	0.84 (0.62, 1.14)	0.25	0.71 (0.57, 0.88)	0.002
rs1532085 (G>A)	40	*LIPC*	+3.0	-1 (-16, +15)	N/Y	0.0	0.95	1.02 (0.86, 1.20)	0.85	1.09 (0.83, 1.42)	0.54	1.01 (0.83, 1.21)	0.96
rs6065906 (T>C)	16	*PLTP/PCIF1*	+3.3	+23 (+2, +43)	Y/Y	4.8	0.03	1.11 (0.90, 1.38)	0.31	0.74 (0.50, 1.08)	0.11	0.91 (0.71, 1.16)	0.44

rs174546 (C>T)	33	*FADS1*	+3.8	+8 (-8, +25)	Y/Y	0.9	0.33	1.15 (0.97, 1.36)	0.12	1.02 (0.78, 1.34)	0.89	1.06 (0.87, 1.28)	0.58
rs2929282 (A>T)	6	*FRMD5*	+5.1	-21 (-56, +14)	N/Y	1.4	0.24	0.89 (0.62, 1.28)	0.52	1.26 (0.76, 2.09)	0.38	1.23 (0.83, 1.82)	0.29
rs2412710 (G>A)	2	*CAPN3*	+7.0	-50 (-109, +9)	N/Y	2.8	0.10	1.33 (0.74, 2.39)	0.34	1.21 (0.48, 3.02)	0.68	1.21 (0.61, 2.40)	0.58
rs1260326 (C>T)	38	*GCKR*	+8.9	+9 (-7, +26)	Y/Y	1.3	0.26	1.02 (0.87, 1.20)	0.80	0.95 (0.73, 1.24)	0.72	1.11 (0.93, 1.34)	0.26
rs964184 (C>G)	**14**	***ZNF259/ APOA1-4-5***	**+17.0**	**+45 (+23, +67)**	**Y/N**	**16.0**	**7×10^{-5}**	**0.98 (0.78, 1.25)**	**0.89**	**1.42 (1.00, 2.00)**	**0.05**	**1.10 (0.85, 1.42)**	**0.49**

[a] SNPs (major allele > minor allele) associated with TG at $P< 5\times10^{-8}$ from GLGC GWAS.

[b] MAF among controls.

[c] mg/dL increase (or decrease) in mean TG per minor allele reported by GLGC GWAS.

[d] mg/dL increase (or decrease) in zenith TG (mg/dL) per minor allele among controls from linear regression model adjusted for age at colonoscopy, sex, race, and data-collection period.

[e] Y if associations from GLGC and controls have same sign before rounding; N otherwise / Y if association from GLGC is contained within the 95% CI from controls before rounding; N otherwise.

[f] OR from polytomous logistic regression model adjusted for age at colonoscopy, sex, race, and data-collection period.

[g] rs4564803 was used as a proxy for rs1042034.

Abbreviations: CI, confidence interval; GLGC, Global Lipids Genetics Consortium; GWAS, genome-wide association study; MAF, minor-allele frequency; N, no; OR, odds ratio; SNP, single-nucleotide polymorphism; TG, triglycerides; Y, yes.

GLGC GWAS results are from Teslovich TM, Musunuru K, Smith AV, et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* 2010;466(7307):707-713.

Supplemental Table 4.6. Genotype-lipid (among controls) and genotype-polyp associations for the single-SNPs associated with TC. SNPs with $P \leq 0.01$ for genotype-lipid association among controls shown in bold.

			TC association	Zenith TC among controls from Group Health				Non-advanced adenomas vs. controls		Advanced adenomas vs. controls		Non-adenomatous polyps vs. controls	
SNP[a]	MAF (%)[b]	Nearest gene(s)	from GLGC GWAS[c]	Association[d] (95% CI)	Direction/ Coverage[e]	*F*	*P*	OR[f] (95% CI)	*P*	OR[f] (95% CI)	*P*	OR[f] (95% CI)	*P*
rs6511720 (G>T)	11	*LDLR*	-7.1	-4 (-12, +3)	Y/Y	1.3	0.26	1.10 (0.86, 1.42)	0.46	0.98 (0.64, 1.49)	0.92	0.80 (0.59, 1.08)	0.15
rs629301 (T>G)	**21**	***CELSR2/ SORT1***	**-5.4**	**-11 (-16, -5)**	**Y/Y**	**12.8**	**4×10^{-4}**	**1.09 (0.90, 1.33)**	**0.38**	**1.07 (0.78, 1.46)**	**0.68**	**1.13 (0.91, 1.40)**	**0.28**
rs10401969 (T>C)	**9**	***SUGP1/CILP2***	**-4.7**	**-11 (-19, -3)**	**Y/Y**	**7.7**	**0.006**	**0.86 (0.65, 1.15)**	**0.31**	**0.83 (0.52, 1.32)**	**0.42**	**0.75 (0.53, 1.05)**	**0.09**
rs1800961 (C>T)	3	*HNF4A*	-4.7	-12 (-25, +1)	Y/Y	3.1	0.08	0.84 (0.52, 1.37)	0.49	1.88 (1.04, 3.40)	0.04	0.76 (0.44, 1.33)	0.33
rs9987289 (G>A)	7	*PPP1R3B*	-3.1	+1 (-8, +10)	N/Y	0.1	0.75	1.04 (0.77, 1.41)	0.78	1.11 (0.69, 1.78)	0.66	1.28 (0.92, 1.78)	0.14
rs2131925 (T>G)	33	*DOCK7/ ANGPTL3*	-2.6	0 (-5, +5)	N/Y	0.0	0.99	0.94 (0.79, 1.12)	0.49	1.15 (0.87, 1.51)	0.32	1.03 (0.85, 1.25)	0.78
rs2954029 (A>T)	46	*TRIB1*	-2.3	-2 (-7, +3)	Y/Y	0.7	0.42	1.03 (0.88, 1.20)	0.75	1.06 (0.82, 1.37)	0.63	1.17 (0.98, 1.40)	0.09
rs1800562 (G>A)	5	*HFE*	-2.2	-9 (-19, +2)	Y/Y	2.4	0.12	1.36 (0.96, 1.92)	0.08	0.97 (0.53, 1.77)	0.91	1.45 (0.99, 2.13)	0.06
rs1883025 (C>T)	27	*ABCA1*	-2.2	-1 (-7, +4)	Y/Y	0.2	0.69	0.91 (0.76, 1.10)	0.34	1.20 (0.90, 1.60)	0.23	1.08 (0.88, 1.34)	0.47
rs6882076 (C>T)	36	*TIMD4*	-2.0	-2 (-7, +3)	Y/Y	0.7	0.42	1.07 (0.90, 1.26)	0.44	1.30 (1.00, 1.69)	0.05	0.91 (0.75, 1.10)	0.33
rs2814982 (C>T)	11	*C6orf106*	-1.9	-6 (-14, +1)	Y/Y	2.6	0.11	1.09 (0.85, 1.41)	0.50	1.10 (0.74, 1.64)	0.64	1.10 (0.82, 1.46)	0.53
rs174546 (C>T)	33	*FADS1*	-1.8	+1 (-4, +6)	N/Y	0.3	0.58	1.15 (0.97, 1.36)	0.12	1.02 (0.78, 1.34)	0.89	1.06 (0.87, 1.28)	0.58
rs581080 (C>G)	19	*TTC39B*	-1.6	+2 (-4, +7)	N/Y	0.3	0.62	0.88 (0.72, 1.09)	0.24	0.71 (0.50, 1.02)	0.06	0.99 (0.79, 1.24)	0.92
rs3757354 (C>T)	24	*MYLIP*	-1.5	-4 (-9, +1)	Y/Y	2.1	0.15	0.86 (0.70, 1.04)	0.12	0.92 (0.67, 1.25)	0.58	0.89 (0.71, 1.11)	0.29
rs2290159 (G>C)	21	*RAF1*	-1.4	-4 (-10, +2)	Y/Y	1.6	0.21	0.93 (0.76, 1.13)	0.45	0.89 (0.65, 1.23)	0.49	1.04 (0.84, 1.29)	0.72
rs2642442 (T>C)	31	*MARC1*	-1.4	-4 (-9, +1)	Y/Y	2.6	0.11	1.07 (0.90, 1.27)	0.44	1.21 (0.93, 1.59)	0.16	1.04 (0.85, 1.26)	0.72
rs2902940 (A>G)	28	*MAFB*	-1.4	-3 (-8, +3)	Y/Y	0.9	0.36	1.06 (0.89, 1.27)	0.53	1.15 (0.86, 1.52)	0.35	1.08 (0.88, 1.32)	0.45
rs514230 (A>T)	45	*IRF2BP2*	-1.4	-2 (-6, +3)	Y/Y	0.4	0.51	1.05 (0.89, 1.23)	0.58	0.92 (0.71, 1.20)	0.54	1.05 (0.87, 1.25)	0.62
rs12027135 (T>A)	48	*TMEM57/ LDLRAP1*	-1.2	+2 (-3, +7)	N/Y	0.7	0.39	0.98 (0.83, 1.15)	0.81	0.94 (0.72, 1.21)	0.61	0.87 (0.73, 1.05)	0.15
rs495565 (A>G)[g]	35	*FRK*	-1.2	-6 (-11, -1)	Y/Y	5.5	0.02	1.08 (0.91, 1.27)	0.39	0.99 (0.75, 1.29)	0.92	1.13 (0.93, 1.36)	0.21
rs989653 (G>A)[h]	21	*EVI5*	-1.2	0 (-5, +6)	N/Y	0.0	0.93	0.87 (0.72, 1.07)	0.19	0.94 (0.68, 1.29)	0.68	1.00 (0.80, 1.24)	0.97
rs2737229 (A>C)	30	*TRPS1*	-1.1	+1 (-5, +6)	N/Y	0.1	0.83	0.96 (0.80, 1.15)	0.65	1.39 (1.05, 1.82)	0.02	0.94 (0.77, 1.15)	0.57
rs10128711 (C>T)	32	*SPTY2D1*	-1.0	+2 (-3, +7)	N/Y	0.6	0.46	0.92 (0.77, 1.09)	0.33	1.02 (0.78, 1.34)	0.87	0.97 (0.80, 1.18)	0.74
rs11065987 (A>G)	40	*BRAP*	-1.0	-3 (-8, +2)	Y/Y	1.8	0.18	0.99 (0.84, 1.17)	0.91	1.18 (0.90, 1.54)	0.23	1.03 (0.86, 1.25)	0.73
rs7206971 (G>A)	46	*EFCAB13*	+1.0	+3 (-2, +7)	Y/Y	1.3	0.26	1.10 (0.94, 1.29)	0.24	1.04 (0.81, 1.35)	0.76	1.20 (1.01, 1.44)	0.04
rs7941030 (T>C)	39	*UBASH3B*	+1.0	+1 (-4, +6)	Y/Y	0.1	0.75	0.99 (0.84, 1.17)	0.94	1.08 (0.83, 1.41)	0.55	0.91 (0.76, 1.10)	0.33

rs1495741 (A>G)	28	*NAT2*	+1.1	+3 (-3, +8)	Y/Y	1.0	0.32	0.78 (0.65, 0.94)	0.01	0.84 (0.62, 1.14)	0.25	0.71 (0.57, 0.88)	0.002
rs2255141 (G>A)	29	*GPAM*	+1.1	-1 (-6, +4)	N/Y	0.1	0.72	1.07 (0.90, 1.27)	0.43	0.97 (0.74, 1.29)	0.85	0.92 (0.76, 1.12)	0.43
rs2081687 (C>T)	32	*CYP7A1*	+1.2	+3 (-2, +8)	Y/Y	1.6	0.21	0.99 (0.83, 1.18)	0.92	0.85 (0.64, 1.13)	0.27	1.07 (0.88, 1.30)	0.50
rs11136341 (A>G)	35	*PLEC*	+1.3	-1 (-6, +4)	N/Y	0.1	0.74	1.06 (0.90, 1.26)	0.49	1.52 (1.17, 1.97)	0.002	1.07 (0.88, 1.29)	0.50
rs492602 (A>G)	47	*FUT2*	+1.3	-4 (-8, +1)	N/N	2.1	0.15	0.91 (0.77, 1.07)	0.26	1.04 (0.80, 1.36)	0.76	1.11 (0.93, 1.34)	0.25
rs7570971 (C>A)	40	*RAB3GAP1*	+1.3	-3 (-8, +2)	Y/Y	1.9	0.17	1.06 (0.89, 1.25)	0.52	1.07 (0.82, 1.39)	0.64	1.00 (0.82, 1.20)	0.96
rs1169288 (A>C)	32	*HNF1A*	+1.4	+2 (-3, +7)	Y/Y	0.6	0.43	1.01 (0.84, 1.21)	0.95	1.10 (0.83, 1.46)	0.51	1.06 (0.86, 1.29)	0.60
rs12670798 (T>C)	23	*DNAH11*	+1.4	+5 (0, +11)	Y/Y	3.5	0.06	1.05 (0.87, 1.28)	0.60	1.56 (1.17, 2.08)	0.003	1.09 (0.87, 1.35)	0.46
rs1532085 (G>A)	40	*LIPC*	+1.5	+4 (-1, +9)	Y/Y	2.8	0.10	1.02 (0.86, 1.20)	0.85	1.09 (0.83, 1.42)	0.54	1.01 (0.83, 1.21)	0.96
rs6029526 (T>A)	50	*TOP1*	+1.5	0 (-5, +5)	N/Y	0.0	1.00	1.02 (0.87, 1.20)	0.80	0.86 (0.67, 1.11)	0.26	1.04 (0.87, 1.24)	0.70
rs3764261 (C>A)	31	*CETP*	+1.7	+3 (-2, +8)	Y/Y	1.6	0.20	0.96 (0.81, 1.15)	0.68	0.98 (0.74, 1.30)	0.89	1.09 (0.90, 1.33)	0.40
rs1260326 (C>T)	38	*GCKR*	+1.9	+2 (-3, +7)	Y/Y	0.6	0.44	1.02 (0.87, 1.20)	0.80	0.95 (0.73, 1.24)	0.72	1.11 (0.93, 1.34)	0.26
rs11220462 (G>A)	14	*ST3GAL4*	+2.0	+2 (-5, +9)	Y/Y	0.4	0.53	0.98 (0.78, 1.25)	0.88	0.96 (0.65, 1.41)	0.84	1.09 (0.84, 1.41)	0.51
rs2072183 (G>C)	25	*NPC1L1*	+2.0	0 (-6, +6)	N/Y	0.0	0.99	1.05 (0.87, 1.27)	0.62	0.79 (0.57, 1.09)	0.15	1.05 (0.85, 1.29)	0.68
rs2479409 (A>G)	35	*PCSK9*	+2.0	-2 (-6, +4)	N/Y	0.3	0.56	0.87 (0.74, 1.04)	0.13	1.10 (0.84, 1.43)	0.50	1.17 (0.96, 1.41)	0.11
rs1564348 (T>C)	15	*SLC22A1/LPA*	+2.2	+2 (-5, +9)	Y/Y	0.4	0.51	0.84 (0.66, 1.07)	0.16	0.87 (0.59, 1.28)	0.47	0.83 (0.64, 1.09)	0.18
rs2000999 (G>A)	18	*HPR*	+2.3	+2 (-4, +8)	Y/Y	0.4	0.52	1.16 (0.96, 1.42)	0.13	1.05 (0.76, 1.44)	0.77	0.94 (0.75, 1.18)	0.58
rs3177928 (G>A)	13	*HLA-DRA*	+2.3	+3 (-4, +10)	Y/Y	0.8	0.38	1.13 (0.89, 1.43)	0.33	1.07 (0.72, 1.57)	0.75	1.08 (0.83, 1.42)	0.56
rs635634 (C>T)[i]	21	*ABO*	+2.3	+2 (-4, +8)	Y/Y	0.3	0.57	0.85 (0.69, 1.04)	0.12	0.88 (0.64, 1.22)	0.45	0.81 (0.65, 1.02)	0.07
rs12916 (T>C)	36	*HMGCR*	+2.8	+5 (0, +10)	Y/Y	4.2	0.04	0.91 (0.77, 1.07)	0.25	1.17 (0.91, 1.51)	0.21	1.08 (0.90, 1.29)	0.40
rs4299376 (T>G)	31	*ABCG8*	+3.0	+5 (0, +10)	Y/Y	3.5	0.06	1.01 (0.85, 1.21)	0.90	0.97 (0.73, 1.29)	0.84	0.87 (0.71, 1.06)	0.16
rs964184 (C>G)	**14**	***ZNF259/ APOA1-4-5***	**+4.7**	**+12 (+6, +19)**	**Y/N**	**13.4**	**3×10^{-4}**	**0.98 (0.78, 1.25)**	**0.89**	**1.42 (1.00, 2.00)**	**0.05**	**1.10 (0.85, 1.42)**	**0.49**
rs4420638 (A>G)	16	*APOE/APOC1*	+6.8	+7 (-2, +15)	Y/Y	2.4	0.12	1.30 (1.01, 1.69)	0.04	1.18 (0.78, 1.79)	0.44	1.22 (0.94, 1.60)	0.14

[a] SNPs (major allele > minor allele) associated with TC at $P< 5\times10^{-8}$ from GLGC GWAS.

[b] MAF among controls.

[c] mg/dL increase (or decrease) in mean TC per minor allele reported by GLGC GWAS.

[d] mg/dL increase (or decrease) in zenith TC (mg/dL) per minor allele among controls from linear regression model adjusted for age at colonoscopy, sex, race, and data-collection period.

[e] Y if associations from GLGC and controls have same sign before rounding; N otherwise / Y if association from GLGC is contained within the 95% CI from controls before rounding; N otherwise.

[f] OR from polytomous logistic regression model adjusted for age at colonoscopy, sex, race, and data-collection period.

[g] rs495565 was used as a proxy for rs9488822.

[h] rs989653 was used as a proxy for rs7515577.

[i] rs635634 replaced rs9411489 according to a change made to the NCBI reference sequence.

[g] rs495565 was used as a proxy for rs9488822.

Abbreviations: CI, confidence interval; GLGC, Global Lipids Genetics Consortium; GWAS, genome-wide association study; MAF, minor-allele frequency; N, no; NCBI, National Center for Biotechnology Information; OR, odds ratio; SNP, single-nucleotide polymorphism; TC, total cholesterol; Y, yes.

GLGC GWAS results are from Teslovich TM, Musunuru K, Smith AV, et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* 2010;466(7307):707-713.

Supplemental Table 4.7. Genotype-lipid associations using trait-specific multi-SNP allele scores, stratified by case-control status.

		Controls (N=754)			Non-advanced adenomas (N=518)			Advanced adenomas (N=139)			Non-adenomatous polyps (N=380)		
Genetic exposure[a]	Adjustment[b]	Association for +1 score[c] (95% CI)	*F*	*P*	Association for +1 score[c] (95% CI)	*F*	*P*	Association for +1 score[c] (95% CI)	*F*	*P*	Association for +1 score[c] (95% CI)	*F*	*P*
LDL Allele Score	Minimal	+2.3 (+1.4, +3.2)	24.6	1×10^{-6}	+2.1 (+1.2, +3.2)	19.5	1×10^{-5}	+2.4 (+0.5, +4.3)	6.3	0.01	+2.6 (+1.4, +3.7)	19.1	2×10^{-5}
	Full	+2.4 (+1.5, +3.4)	25.7	6×10^{-7}	+2.2 (+1.1, +3.3)	16.2	7×10^{-5}	+2.7 (0.0, +5.3)	3.8	0.06	+2.6 (+1.3, +3.8)	16.0	9×10^{-5}
HDL Allele Score	Minimal	-0.9 (-1.1, -0.7)	77.3	1×10^{-17}	-0.9 (-1.1, -0.6)	45.4	4×10^{-11}	-0.9 (-1.4, -0.4)	11.6	9×10^{-4}	-0.7 (-1.0, -0.4)	20.1	1×10^{-5}
	Full	-0.9 (-1.1, -0.7)	72.9	9×10^{-17}	-0.8 (-1.1, -0.6)	49.8	7×10^{-12}	-1.0 (-1.6, -0.4)	11.2	1×10^{-3}	-0.7 (-1.0, -0.4)	23.0	2×10^{-6}
TG Allele Score	Minimal	+9.8 (+6.6, +12.9)	37.4	2×10^{-9}	+10.3 (+6.5, +14.1)	28.2	2×10^{-7}	+14.6 (+6.7, +22.4)	13.2	5×10^{-4}	+9.7 (+5.6, +13.9)	21.1	7×10^{-6}
	Full	+10.1 (+7.0, +13.2)	41.0	4×10^{-10}	+10.3 (+6.5, +14.2)	27.7	3×10^{-7}	+18.0 (+8.9, +27.0)	15.1	3×10^{-4}	+8.2 (+4.0, +12.3)	14.7	2×10^{-4}
TC Allele Score	Minimal	+2.5 (+1.8, +3.3)	44.6	5×10^{-11}	+2.5 (+1.4, +3.5)	22.2	3×10^{-6}	+4.6 (+2.6, +6.6)	19.9	2×10^{-5}	+2.8 (+1.9, +3.8)	32.4	3×10^{-8}
	Full	+2.4 (+1.7, +3.2)	39.4	6×10^{-10}	+2.3 (+1.2, +3.4)	17.9	3×10^{-5}	+3.3 (+0.9, +5.7)	7.2	9×10^{-3}	+2.8 (+1.8, +3.8)	31.5	4×10^{-8}

[a] Linear regression model with highest lipid measurement (lowest for HDL) in 20 years prior to colonoscopy as the dependent variable and count of alleles associated with increasing lipid measurement (decreasing for HDL) in the GLGC GWAS weighted by the effect size from GLGC GWAS as the independent variable.

[b] Minimal estimate adjusted for age at colonoscopy, sex, race, and data-collection period. Full estimate adjusted for age at colonoscopy, sex, race, education, BMI, NSAID use, first-degree family history of colorectal cancer, estrogen-only use (in women), estrogen-plus-progestin use (in women), cigarette smoking, alcohol consumption, diabetes mellitus, fruit servings per day, vegetable servings per day, recreational and exercise physical activity, prior endoscopy (approximately 2 years before study colonoscopy), and data-collection period.

[c] mg/dL increase (or decrease) in mean lipid trait per unit increase in score. Lipid component-specific allele score provides a count of alleles associated with increasing lipid measurement (decreasing for HDL) in the GLGC GWAS, weighted by the effect size from GLGC GWAS.

Abbreviations: BMI, body mass index; CI, confidence interval; GLGC, Global Lipids Genetics Consortium; GWAS, genome-wide association study; HDL, high-density lipoprotein cholesterol; LDL, low-density lipoprotein cholesterol; NSAID, nonsteroidal anti-inflammatory drug; SNP, single-nucleotide polymorphism; TC, total cholesterol; TG triglycerides.

GLGC GWAS results are from Teslovich TM, Musunuru K, Smith AV, et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* 2010;466(7307):707-713.

Supplemental Table 4.8. Association between GLGC-weighted allele scores and risk factors for colorectal polyps, including pre-colonoscopy lipid phenotypes. The value of β represents the difference in mean allele score relative to the reference category.

	LDL Allele Score			HDL Allele Score			TG Allele Score			TC Allele Score		
Variable	β^a	SE	*P*	β^a	SE	*P*	β^a	SE	*P*	β^a	SE	*P*
Education	Ref		0.98	Ref		0.77	Ref		0.69	Ref		0.98
High school or less	-0.02	0.31		+0.16	0.36		-0.27	0.29		-0.01	0.34	
Some college	-0.06	0.31		-0.01	0.35		-0.34	0.28		+0.08	0.34	
College	+0.03	0.30		+0.25	0.34		-0.26	0.27		-0.03	0.33	
Graduate												
Body mass index (BMI), kg/m^2			0.07			0.53			0.56			0.15
<25	Ref			Ref			Ref			Ref		
25-30	+0.48	0.22		+0.25	0.25		+0.03	0.20		+0.46	0.24	
≥30	+0.40	0.25		+0.26	0.29		+0.23	0.23		+0.32	0.28	
Family history of CRC in first-degree relatives			0.61			0.41			0.99			0.75
No	Ref			Ref			Ref			Ref		
Yes	+0.12	0.23		+0.21	0.26		+0.00	0.21		+0.08	0.25	
Diabetes mellitus			0.96			0.46			0.85			0.74
No	Ref			Ref			Ref			Ref		
Yes	-0.02	0.33		-0.28	0.37		+0.06	0.30		+0.12	0.36	
NSAID use			0.47			0.21			0.93			0.71
Never	Ref			Ref			Ref			Ref		
Former	-0.16	0.35		-0.34	0.39		-0.04	0.32		-0.10	0.38	
Current	+0.19	0.20		-0.39	0.23		+0.06	0.19		+0.15	0.22	
Estrogen-alone use			0.96			0.52			0.29			0.87
No	Ref			Ref			Ref			Ref		
Yes	-0.01	0.27		+0.20	0.31		+0.26	0.25		+0.05	0.30	
Estrogen-plus-progestin use			0.41			0.40			0.67			0.34
No	Ref			Ref			Ref			Ref		
Yes	-0.23	0.28		+0.27	0.32		-0.11	0.25		-0.29	0.30	
Cigarette smoking			0.16			0.03			0.21			0.09

Never	Ref			Ref			Ref			Ref		
Former	+0.36	0.20		-0.58	0.22		-0.31	0.18		+0.35	0.22	
Current	-0.07	0.39		-0.05	0.44		-0.01	0.36		-0.47	0.43	
Alcohol consumption, drinks/week			0.59			0.43			0.01			0.14
0	Ref			Ref			Ref			Ref		
>0-<7	+0.68	0.22		-0.23	0.25		-0.23	0.20		+0.08	0.24	
7-14	-0.27	0.31		-0.32	0.34		-0.44	0.28		-0.60	0.33	
≥14	-0.29	0.32		-0.55	0.36		-0.94	0.29		-0.46	0.35	
Vegetable consumption, servings/day			0.31			0.63			0.89			0.14
0-<1	Ref			Ref			Ref			Ref	0.29	
1-2	-0.26	0.27		-0.06	0.31		-0.04	0.25		-0.42	0.29	
2-3	+0.19	0.28		-0.17	0.32		-0.11	0.25		+0.18	0.31	
≥3	-0.10	0.31		+0.24	0.35		+0.09	0.28		-0.08	0.34	
Fruit consumption, servings/day			0.89			0.17			0.25			0.77
0-<1	Ref			Ref			Ref			Ref		
1-2	+0.19	0.25		+0.47	0.28		+0.36	0.23		+0.12	0.27	
2-3	+0.06	0.26		+0.57	0.30		+0.32	0.24		-0.05	0.29	
≥3	+0.11	0.30		+0.62	0.34		-0.01	0.27		+0.27	0.33	
Recreational and exercise physical activity, hours/week			0.83			0.54			0.20			0.96
0	Ref			Ref			Ref			Ref		
>0-<1	+0.28	0.38		+0.55	0.43		+0.39	0.35		+0.18	0.42	
1-2	+0.07	0.35		+0.30	0.40		-0.13	0.32		-0.03	0.38	
2-6	+0.09	0.38		+0.67	0.42		+0.22	0.34		-0.09	0.41	
≥6	+0.30	0.36		+0.45	0.40		-0.14	0.32		+0.03	0.39	
Prior lower endoscopy[b]			0.40			0.70			0.21			0.68
No	Ref			Ref			Ref			Ref		
Yes	+0.17	0.20		-0.09	0.22		-0.23	0.18		+0.09	0.22	
Lipid-controlling drug use			9×10^{-8}			2×10^{-3}			2×10^{-10}			6×10^{-8}
No	Ref			Ref			Ref			Ref		
Yes	+1.16	0.22		+0.78	0.25		+1.25	0.20		+1.29	0.24	

Zenith LDL (ATP III)			3×10^{-14}			0.09			2×10^{-4}			2×10^{-13}
Optimal (<100 mg/dL)	Ref			Ref			Ref			Ref		
Near optimal (100-129 mg/dL)	+0.64	0.38		+0.55	0.45		+0.80	0.35		+0.69	0.42	
Borderline high (130-159 mg/dL)	+1.19	0.37		+0.99	0.43		+1.34	0.34		+1.40	0.41	
High (160-189 mg/dL)	+1.97	0.39		+0.95	0.45		+1.45	0.36		+2.09	0.42	
Very high (≥190 mg/dL)	+3.00	0.42		+1.17	0.49		+1.45	0.39		+3.23	0.46	
Nadir HDL (ATP III)			3×10^{-5}			1×10^{-25}			2×10^{-8}			0.08
High (≥60 mg/dL)	Ref			Ref			Ref			Ref		
Optimal (40-59 mg/dL)	-0.29	0.23		-1.57	0.25		-0.78	0.20		+0.00	0.25	
Low (<40 mg/dL)	-1.24	0.28		-3.34	0.31		-1.52	0.56		-0.58	0.31	
Zenith TG (ATP III)			5×10^{-5}			3×10^{-6}			1×10^{-23}			9×10^{-8}
Normal (<150 mg/dL)	Ref			Ref			Ref			Ref		
Borderline high (150-199 mg/dL)	+0.59	0.30		+0.21	0.34		+1.12	0.27		+1.00	0.33	
High (200-499 mg/dL)	+1.02	0.25		+0.83	0.29		+1.71	0.22		+1.39	0.28	
Very high (≥500 mg/dL)	+1.69	0.51		+2.81	0.58		+3.85	0.45		+2.24	0.55	
Zenith TC (ATP III)			1×10^{-16}			0.09			8×10^{-12}			1×10^{-22}
Desirable (<200 mg/dL)	Ref			Ref			Ref			Ref		
Borderline high (200-239 mg/dL)	+0.60	0.28		-0.36	0.32		+0.15	0.25		+0.87	0.30	
High (≥240 mg/dL)	+1.96	0.27		+0.16	0.30		+1.32	0.24		+2.55	0.29	

Abbreviations: ATP, Adult Treatment Panel; CRC, colorectal cancer; GLGC, Global Lipids Genetics Consortium; HDL, high-density lipoprotein cholesterol; LDL, low-density lipoprotein cholesterol; NSAID, nonsteroidal anti-inflammatory drug; Ref, reference category; SE, standard error; TG, triglycerides; TC, total cholesterol.

[a] Adjusted for age at colonoscopy, sex, race, and data-collection period.

[b] Lower endoscopy earlier than approximately 2 years before the study colonoscopy.

Supplemental Table 4.9. Sensitivity analyses for LDL genotype-polyp associations accounting for use of lipid-controlling drugs.

Phenotype[a]	Adjustment[b]	Unit change for OR (+20 mg/dL)	Non-advanced adenomas vs. controls		Advanced adenomas vs. controls		Non-adenomatous polyps vs. controls	
			OR (95% CI)	*P*	OR (95% CI)	*P*	OR (95% CI)	*P*
All lipid-controlling drug users excluded[c]								
LDL Allele Score	Minimal	+18.2 score	1.07 (0.57, 1.99)	0.84	1.67 (0.59, 4.69)	0.33	1.80 (0.91, 3.56)	0.09
LDL Allele Score	Full	+12.5 score	1.00 (0.65, 1.60)	0.99	1.21 (0.58, 2.54)	0.61	1.46 (0.89, 2.39)	0.13
Lipid-controlling drug users with pre-colonoscopy zenith LDL<130 mg/dL excluded[d]								
LDL Allele Score	Minimal	+8.7 score	0.98 (0.76, 1.26)	0.86	1.25 (0.82, 1.90)	0.31	1.11 (0.83, 1.47)	0.49
LDL Allele Score	Full	+8.7 score	1.01 (0.77, 1.33)	0.95	1.17 (0.75, 1.82)	0.48	1.14 (0.84, 1.53)	0.41
Lipid-controlling drug users with pre-colonoscopy zenith LDL<130 mg/dL imputed[e]								
LDL Allele Score	Minimal	+9.1 score	0.99 (0.76, 1.29)	0.95	1.17 (0.77, 1.84)	0.46	1.14 (0.85, 1.51)	0.40
LDL Allele Score	Full	+8.7 score	1.03 (0.79, 1.34)	0.84	1.11 (0.73, 1.70)	0.63	1.16 (0.87, 1.55)	0.32

[a] Polytomous logistic regression model with case-control status as the dependent variable and the highest LDL measurement in at most 20 years prior to colonoscopy as the independent variable.

[b] Minimally-adjusted OR is adjusted for age at colonoscopy, sex, race, and data-collection period. Fully-adjusted OR is adjusted for age at colonoscopy, sex, race, education, BMI, NSAID use, family history of colorectal cancer, estrogen-only use (in women), estrogen-plus-progestin use (in women), cigarette smoking, alcohol consumption, diabetes mellitus, fruit servings per day, vegetable servings per day, recreational and exercise physical activity, prior endoscopy (approximately 2 years before study colonoscopy), and data-collection period.

[c] N=317 controls, N=224 non-advanced adenoma cases, N=60 advanced adenoma cases, N=171 non-adenomatous polyp cases.

[d] N=477 controls, N=362 non-advanced adenoma cases, N=94 advanced adenoma cases, N=252 non-adenomatous polyp cases.

[e] N=513 controls, N=390 non-advanced adenoma cases, N=103 advanced adenoma cases, N=269 non-adenomatous polyp cases. Single-imputed value = max(100 mg/dL, observed zenith LDL + 30 mg/dL).

Abbreviations: BMI, body mass index; CI, confidence interval; LDL, low-density lipoprotein cholesterol; NSAID, nonsteroidal anti-inflammatory drug; OR, odds ratio.

5. Conclusions and future directions

Summary of Chapters 1-4

We observed associations between pre-colonoscopy blood cholesterol levels and the prevalence of colorectal adenomatous polyps (Chapter 4). In particular, high LDL, high TG, and, low HDL, as measured by their peak over time, were associated with increased prevalence of advanced adenomas, those most likely to progress to adenocarcinoma. These results are generally consistent with the average effects observed from endoscopy studies that assessed blood cholesterol concentrations from a fasting blood draw at the time of endoscopy (Chapter 2).

We found that self-report of hypercholesterolemia was generally accurate, with higher specificity than sensitivity (Chapter 3), and those who reported untreated hypercholesterolemia were more likely to have prevalent adenomas compared to those who reported treated hypercholesterolemia (Chapter 1). Consistently, receipt of statin therapy was associated with lower odds of having adenomas, only among, those who had LDL levels >160 mg/dL at any time prior to the study colonoscopy (Chapter 1).

Using single-nucleotide polymorphisms (SNPs) identified by the Global Lipids Genetics Consortium to be associated with blood cholesterol levels as instrumental variables in a Mendelian randomization study, there was insufficient evidence to conclude that genetic susceptibility to dyslipidemia is associated with colorectal polyps (Chapter 4). In this light, observed associations between blood lipid concentrations and colorectal polyps may be non-causal, and dyslipidemia may be a marker of the type of visceral adiposity and dietary exposures that promote neoplastic growth in the colon and rectum. On the other hand, the limited statistical power of our Mendelian randomization analysis may account for our inability to observe any genotype-polyp associations.

Limitations

We recapitulate 3 main limitations of this research, and discuss recommendations for future studies to address each. These limitations are: 1) lack of statistical power; 2) imperfect assessment of timing of relevant cholesterol exposures; and 3) blood lipid concentrations may not adequately measure cholesterol functionality.

Limitation 1: Lack of statistical power

As discussed in Chapter 4, the statistical power of the Mendelian randomization approach is limited by both the strength of the instrument and its variance explained (187). The SNPs identified by the Global Lipids Genetics Consortium collectively account for ~12% of the total variation in observed cholesterol levels (12% for LDL, 12% for HDL, 12% for TG, and 10% for TC) (41). Although ~12% is relatively modest, this magnitude of variance explained is among the highest identified for common human traits (e.g., for BMI and blood pressure, for example, the amount of variance explained by known common loci is estimated to be <5%) (188).

In our analyses of cholesterol phenotypes, we conducted exploratory subgroup analyses, including examining heterogeneity by sex, prior endoscopy status, the magnitude of achieved zenith LDL, and other factors. Such stratification in our analyses of genotypes would be extremely underpowered, given the sample size requirements for Mendelian randomization. In Chapter 4, for example, we explored associations by anatomic location, primarily motivated by the bile acid hypothesis, which posits that those with hypercholesterolemia may have higher levels of residual bile acids in the large intestine which serve as a source of chronic inflammation leading to lesions in the proximal colon where the bile acids may be more potent.

Addressing Limitation 1: Larger studies

In our study, there was no concordance between which SNPs were associated with lipid phenotypes and which were associated with polyps. In general, the SNPs with the largest magnitude per-allele associations with lipid phenotypes were not associated with colorectal polyps. Despite the lack of evidence of we found for a genotype-polyp association, we feel that similar Mendelian randomization analyses should be conducted in larger studies.

A readily available opportunity is to perform an analysis using the Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO) (185). Although this collaborative effort is primarily focused on colorectal cancer, and not polyps, there are a few participating studies that contribute data on adenoma cases, including the Tennessee Colorectal Polyp Study, Nurses' Health Study, Health Professionals' Follow-up Study, and the Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial. Although statistical power for Mendelian randomization analyses will be higher with a larger sample size, GECCO has limited ability to harmonize pathology information, past lipid trajectories, and pharmacy data.

Limitation 2: Imperfect assessment of timing of relevant cholesterol exposures

Both hypercholesterolemia and colorectal polyps likely develop for several years before being clinically detectable, making it difficult to establish the temporality of exposure and outcome. Even if we are confident that the onset of hypercholesterolemia precedes the development of colorectal lesions among all study participants, it is unclear to what magnitude or how long before polyp occurrence hypercholesterolemia needs to be present in order to be considered etiologically relevant. Hypercholesterolemia may have latency. Having ever had

abnormal cholesterol, even if later having returned to lower-risk levels as a consequence of treatment, may nonetheless impart persistent higher long-term risk compared to those who never leave a low-risk state. That is, just as "former smokers" retain some elevated risk of many chronic diseases compared to "never smokers", "former hypercholesterolemics" may retain elevated risk of developing colorectal neoplasia compared to "never hypercholesterolemics".

In Chapters 3 and 4, we used the zenith pre-colonoscopy LDL, TG, and TC (nadir HDL) as the main measurements of cholesterol exposure. These "high-water marks" were intended to represent the magnitude of hypercholesterolemia, prior to initiation of lipid-controlling therapy. This approach was motivated by studies of prostate cancer that utilize nadir PSA over follow-up (189). It may, however, be important to consider the timing of peak hypercholesterolemia, as the risk of developing colorectal polyps may increase only after hypercholesterolemia has been present for several years. We could use data collected for this dissertation to test associations between polyp prevalence and age at zenith LDL.

Nationwide, the use of statins has greatly increased during the time this study was conducted (190). As statins are more broadly recommended, therapy may be initiated at lower LDL levels. The true latent LDL level may be unobserved and the zenith LDL is a censored version with censoring value decreasing over time (see areas for future methodological development below). On the other hand lipid-controlling drugs, including statins, have been shown to have anti-proliferative and anti-inflammatory properties *in vitro* (191), and long-term use may reduce the risk of colorectal neoplasia independent of lipid-lowering effects. Statins, however, are usually only prescribed to those with hypercholesterolemia (i.e., off-label uses are extremely uncommon), so it is difficult to study the effect of statins on reducing the risk of colorectal neoplasia among those who do not have hypercholesterolemia.

Addressing Limitation 2: Assess colorectal polyp incidence longitudinally

This dissertation retrospectively incorporated longitudinal exposure information. Polyp outcome, however, was determined during only a cross-sectional assessment (i.e., the study colonoscopy). Results provided in Chapter 1 suggest that our associations of interest may differ among those who did or did not have a previous endoscopy. Indeed, controls in our study may have once had a polyp detected and removed before the start of the study. The contamination of our control group with some proportion of polyp-formers may have reduced our statistical power to detect phenotype-polyp and genotype-polyp associations.

In our evaluation based on medical records, some participants had their cholesterol measured more frequently than others, and those with more measurements may have had more opportunities to be to be diagnosed with hypercholesterolemia. An improved design would assess both polyp incidence and cholesterol levels longitudinally, for example, in the context of a cohort study with serial prospective colonoscopies and blood cholesterol measurements at pre-defined intervals over the course of follow-up. Cohort studies would also permit the assessment of competing risks. Because cardiovascular disease and cancer are the two main causes of morbidity and mortality in older adults, exposures that limit increases in cardiovascular disease risk, such as dietary modification or statin therapy, may appear to increase cancer risk (192).

The cohort study design permits the modelling of time-dependent relative risks. Secular trends in the definition and management of hypercholesterolemia may have influenced our results. Increasingly aggressive screening protocols for primary and secondary cardiovascular disease prevention has altered the distribution of cholesterol phenotypes observed in clinical practice. This is an important motivation for using genotypes as a proxy. Allele frequencies for

variants known to be associated with cholesterol levels have been largely unchanged for many generations (193).

Chapter 3 included sensitivity analyses that stratified our assessment of the validation of self-reported hypercholesterolemia by the duration of retrospective follow-up (pre-colonoscopy) available from the electronic laboratory records. This type of sensitivity analysis may also be informative for the phenotype and genotype assessments covered in Chapter 4. We attempted to control for secular trends by adjusting for data collection phase (1998-2003, 2004-2007), but more careful modeling of the time-dependence of variables may be necessary to validly assess whether our associations of interest are dependent on shifts in clinical practice.

Our analyses using the zenith or nadir do not take full advantage of the longitudinal nature of the repeated cholesterol measurements. Calendar-time or age-specific means may be useful to consider in future analyses, along with slopes over time. Zenith and nadir are order statistics and take into account the rank of other measurements, but these statistics do not have variances as small as that of a mean or mean regression function. Some caution is needed, however, when using slopes from retrospective data, as evidenced by the recent controversies surrounding the inability of prospective data to replicate associations between prostate-specific antigen (PSA) velocity and prostate cancer progression measured retrospectively (194, 195).

Limitation 3: Blood lipid concentrations may not adequately measure cholesterol functionality

It has been recently recognized that blood lipid phenotypes, particularly HDL, may have limited ability to capture some information on biological mechanisms reflective of the size,

density, charge, and functionality of apolipoprotein particles (196). Standard measurement of HDL from plasma is primarily indicative of the concentration of large, uncondensed, lipid-rich HDL particles (197), and emerging evidence suggests that measures of functionality may be informative about cardiovascular disease risk beyond values of traditional blood cholesterol fractions (198).

Genotypes identified to be associated with blood lipid phenotypes, are likely to be similarly limited as surrogates of cholesterol functionality. The Global Lipid Genetics Consortium (GLGC) identified common variants based on a cross-sectional cholesterol measurement, without considering functionality. Common SNPs of *PON1*, for example, known related to HDL functionality (199), are not among those identified by the GLGC GWAS. Only variants associated with LDL cholesterol – none of the SNPs linked to HDL or TG – were found to be associated with risk of coronary artery disease (200). Likewise, a recent Mendelian randomization study of myocardial infarction (MI) revealed that variants related to LDL appear to be associated with risk of MI, but genes related to HDL are not (201). It is possible that the lack of association observed in genotypes related to HDL may result from the relatively poor ability of the blood measurement to capture more subtle features such as size and functionality.

Addressing Limitation 3: Measurement of cholesterol functionality

Even in the absence of associations between cholesterol genotypes and colorectal polyp occurrence, future studies may be warranted that evaluate measurements of cholesterol functionality. These mechanisms include: 1) cholesterol efflux capacity (198, 202), as measured by radiolabelled ^{3}H-cholesterol from macrophages (203); 2) anti-inflammatory functionality, as measured by various inflammatory biomarkers from blood or tissue, including interleukins (e.g.,

IL-6) and C-reactive protein (204); and 3) antioxidative functionality, as measured by artlesterase and paraoxonase activity (199). Other biomarkers that may warrant study include 27-hydroxycholesterol, a cholesterol metabolite and ligand to the estrogen receptor, which was recently linked to hormonal cancers (205), and serum amyloid A, known to be enriched in dysfunctional HDL and observed to be overexpressed in colorectal cancer cell-lines (206). It should be noted, however, that if cholesterol dysfunction is casually related to the development of colorectal neoplasia, the magnitude of the association is likely smaller than that between cholesterol dysfunction and cardiovascular disease -- likely far too small to detect in our study sample.

Areas for future methodological development

In conducting this research, we encountered a number of statistical issues that, to our knowledge, have not been fully considered in the methodological literature.

Meta-regression approach

Our primary analyses of Chapter 4 involved the use of a weighted allele score. As a secondary analysis, without using an allele score, we regressed per-allele log-ORs on the association from the GLGC GWAS and tested for the statistical significance of slopes using inverse-variance-weighted linear regression. This meta-regression approach for Mendelian randomization analyses has not been well-described (159). An evaluation of the statistical power of this method, in comparison to the allele score method, is needed.

Multiple comparison adjustment for assessment of instrument strength

Assessing the strength of instrumental variables is also challenging with multiple instruments. F-statistics have been traditionally used to assess instrument strength in instrumental variables analyses (160). $F>10$ is generally taken to indicate a strong instrument. When multiple SNPs are used as instruments in a Mendelian randomization analysis, a multiple comparison correction is necessary. Given the strong assumptions of instrumental variables analysis, it is not clear if standard approaches for multiple testing (e.g., Bonferroni or false discovery rate (207)) are appropriate in this context. This issue warrants further consideration in the statistical literature.

Quantitative trait adjustment to account for treatment

Individuals treated with lipid-controlling drugs typically have their cholesterol values plateau to a treatment-controlled steady state (treated values have both smaller mean and smaller variance than untreated values). Adding a fixed or random treatment effect to each treated measurement is one method to estimate a counterfactual untreated trajectory (161). The true zenith may be unobserved, but a pre-treatment zenith is likely to be closer to the true zenith than the zenith of post-treatment values. Imputation approaches, such as censored normal regression, have been suggested to deal with this potential for bias (208), but further methodological development is warranted.

In Chapter 4, we also noted the possibility that for some participants using lipid-controlling drugs, only post-treatment blood lipid values may be available from Group Health records, and the true zenith LDL may be higher than observed. To account for this, we performed a sensitivity analysis that single-imputed plausible values under various assumptions. This issue is perhaps best exemplified by genetic association studies of commonly treated

cardiovascular disease risk factors, such as hypertension and cholesterol. Genome-wide association studies (GWAS) of blood pressure phenotypes have applied a "correction factor" to the observed blood pressure value among participants taking antihypertensive drugs in an attempt to recover the underlying latent (unobserved) pre-treatment value (209). Similar approaches have been undertaken for GWAS of cholesterol phenotypes for participants taking lipid-controlling drugs (41, 200).

These methods have received increased attention in the literature, having been included as one of the recommendations of the STrengthening the REporting of Genetic Association studies (STREGA) statement (210). Closely following the format of the 22-item STrengthening the Reporting of OBservational Studies in Epidemiology (STROBE) checklist (211), the STREGA statement describes 12 extensions for genetic association studies including recommendations for the reporting of methods for genotyping, polymorphism selection, genotype imputation, control for population stratification, and multiple statistical comparisons. Tobin et al. summarized 10 different approaches for quantitative trait adjustment to account for treatment in genetic association studies: 1) no adjustment; 2) exclude treated; 3) binary treatment variable; 4) binary trait substitution; 5) fixed substitution; 6) random substitution; 7) median substitution; 8) fixed treatment effect; 9) residual adjustment; 10) censored normal regression (161, 212).

Because we collected longitudinal pre-colonoscopy data including pre-treatment values for nearly all participants using lipid-controlling drugs, our analysis did not need to heavily rely on these methods. There was a small subgroup of participants that used lipid-controlling drugs and had pre-colonoscopy zenith LDL<130 mg/dL. Although this situation is plausible given clinical recommendations, we explored how our results would change if we single-imputed

higher zenith LDL values for these individuals. The strategy we chose to implement, however, is a slightly modified version of fixed substitution as described by Tobin et al. based on outside knowledge. For users with pre-colonoscopy zenith LDL<130 mg/dL, the zenith LDL was imputed as the maximum of 100 mg/dL or 30 mg/dL higher than the observed zenith LDL. We selected 100 mg/dL as a floor based on ATP III guidelines and 30 mg/dL as a conservative estimate for the mean 1-year treatment effect of statins from randomized controlled trials (152). It is unclear how this modified approach, setting a minimum based on information from clinical recommendations, performs compared to other potential methods.

References

1. McKeown-Eyssen G. Epidemiology of colorectal cancer revisited: are serum triglycerides and/or plasma glucose associated with risk? *Cancer Epidemiol Biomarkers Prev* 1994;3(8):687-695.

2. Kang HW, Kim D, Kim HJ, et al. Visceral obesity and insulin resistance as risk factors for colorectal adenoma: a cross-sectional, case-control study. *Am J Gastroenterol* 2010;105(1):178-187.

3. Lee GE, Park HS, Yun KE, et al. Association between BMI and metabolic syndrome and adenomatous colonic polyps in Korean men. *Obesity (Silver Spring)* 2008;16(6):1434-1439.

4. Liu CS, Hsu HS, Li CI, et al. Central obesity and atherogenic dyslipidemia in metabolic syndrome are associated with increased risk for colorectal adenoma in a Chinese population. *BMC Gastroenterol* 2010;10:51.

5. Manus B, Adang RP, Ambergen AW, et al. The risk factor profile of recto-sigmoid adenomas: a prospective screening study of 665 patients in a clinical rehabilitation centre. *Eur J Cancer Prev* 1997;6(1):38-43.

6. Kono S, Ikeda N, Yanai F, et al. Serum lipids and colorectal adenoma among male self-defence officials in northern Kyushu, Japan. *Int J Epidemiol* 1990;19(2):274-278.

7. Kono S, Imanishi K, Shinchi K, et al. Serum lipids and left-sided adenomas of the large bowel: an extended study of self-defense officials in Japan. *Cancer Causes Control* 1993;4(2):117-121.

8. Park SK, Joo JS, Kim DH, et al. Association of serum lipids and glucose with the risk of colorectal adenomatous polyp in men: a case-control study in Korea. *J Korean Med Sci* 2000;15(6):690-695.

9. Bird CL, Ingles SA, Frankl HD, et al. Serum lipids and adenomas of the left colon and rectum. *Cancer Epidemiol Biomarkers Prev* 1996;5(8):607-612.

10. Bayerdorffer E, Mannes GA, Ochsenkuhn T, et al. Increased risk of 'high-risk' colorectal adenomas in overweight men. *Gastroenterology* 1993;104(1):137-144.

11. Bayerdorffer E, Mannes GA, Richter WO, et al. Decreased high-density lipoprotein cholesterol and increased low-density cholesterol levels in patients with colorectal adenomas. *Ann Intern Med* 1993;118(7):481-487.

12. Neugut AI, Johnsen CM, Fink DJ. Serum cholesterol levels in adenomatous polyps and cancer of the colon. A case-control study. *JAMA* 1986;255(3):365-367.

13. Rotondo D, Davidson J. Genetics and molecular biology: identification of dyslipidemia genes. *Curr Opin Lipidol* 2010;21(6):548-549.

14. Stampfer MJ, Sacks FM, Salvini S, et al. A prospective study of cholesterol, apolipoproteins, and the risk of myocardial infarction. *N Engl J Med* 1991;325(6):373-381.

15. Poli G, Leonarduzzi G, Biasi F, et al. Oxidative stress and cell signalling. *Curr Med Chem* 2004;11(9):1163-1182.

16. Leonarduzzi G, Arkan MC, Basaga H, et al. Lipid oxidation products in cell signaling. *Free Radic Biol Med* 2000;28(9):1370-1378.

17. Hackman A, Abe Y, Insull W, Jr., et al. Levels of soluble cell adhesion molecules in patients with dyslipidemia. *Circulation* 1996;93(7):1334-1338.

18. Ushio-Fukai M, Hilenski L, Santanam N, et al. Cholesterol depletion inhibits epidermal growth factor receptor transactivation by angiotensin II in vascular smooth muscle cells: role of cholesterol-rich microdomains and focal adhesions in angiotensin II signaling. *J Biol Chem* 2001;276(51):48269-48275.

19. Celletti FL, Waugh JM, Amabile PG, et al. Vascular endothelial growth factor enhances atherosclerotic plaque progression. *Nat Med* 2001;7(4):425-429.

20. English D, Brindley DN, Spiegel S, et al. Lipid mediators of angiogenesis and the signalling pathways they initiate. *Biochim Biophys Acta* 2002;1582(1-3):228-239.

21. Urbich C, Dernbach E, Zeiher AM, et al. Double-edged role of statins in angiogenesis signaling. *Circ Res* 2002;90(6):737-744.

22. Tabas I. Apoptosis and plaque destabilization in atherosclerosis: the role of macrophage apoptosis induced by cholesterol. *Cell Death Differ* 2004;11 Suppl 1:S12-16.

23. Yin J, Chaufour X, McLachlan C, et al. Apoptosis of vascular smooth muscle cells induced by cholesterol and its oxides in vitro and in vivo. *Atherosclerosis* 2000;148(2):365-374.

24. Lawrence T, Willoughby DA, Gilroy DW. Anti-inflammatory lipid mediators and insights into the resolution of inflammation. *Nat Rev Immunol* 2002;2(10):787-795.

25. Dykstra M, Cherukuri A, Sohn HW, et al. Location is everything: lipid rafts and immune cell signaling. *Annu Rev Immunol* 2003;21:457-481.

26. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell* 2011;144(5):646-674.

27. Janakiram NB, Mohammed A, Rao CV. Role of lipoxins, resolvins, and other bioactive lipids in colon and pancreatic cancer. *Cancer Metastasis Rev* 2011.

28. Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* 2003;32(1):1-22.

29. Giovannucci E. Metabolic syndrome, hyperinsulinemia, and colon cancer: a review. *Am J Clin Nutr* 2007;86(3):s836-842.

30. Ebrahim S, Davey Smith G. Mendelian randomization: can genetic epidemiology help redress the failures of observational epidemiology? *Hum Genet* 2008;123(1):15-33.

31. Sheehan NA, Didelez V, Burton PR, et al. Mendelian randomisation and causal inference in observational epidemiology. *PLoS Med* 2008;5(8):e177.

32. Smith GD, Ebrahim S. Mendelian randomization: prospects, potentials, and limitations. *Int J Epidemiol* 2004;33(1):30-42.

33. Cohen JC, Boerwinkle E, Mosley TH, Jr., et al. Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. *N Engl J Med* 2006;354(12):1264-1272.

34. Bennet AM, Di Angelantonio E, Ye Z, et al. Association of apolipoprotein E genotypes with lipid levels and coronary risk. *JAMA* 2007;298(11):1300-1311.

35. Kervinen K, Sodervik H, Makela J, et al. Is the development of adenoma and carcinoma in proximal colon related to apolipoprotein E phenotype? *Gastroenterology* 1996;110(6):1785-1790.

36. Kato I, Land S, Majumdar AP, et al. Functional polymorphisms to modulate luminal lipid exposure and risk of colorectal cancer. *Cancer Epidemiol* 2010;34(3):291-297.

37. Mrkonjic M, Chappell E, Pethe VV, et al. Association of apolipoprotein E polymorphisms and dietary factors in colorectal cancer. *Br J Cancer* 2009;100(12):1966-1974.

38. Reisher SR, Hughes TE, Ordovas JM, et al. Increased expression of apolipoprotein genes accompanies differentiation in the intestinal cell line Caco-2. *Proc Natl Acad Sci U S A* 1993;90(12):5757-5761.

39. Shinomiya S, Sasaki J, Kiyohara C, et al. Apolipoprotein E genotype, serum lipids, and colorectal adenomas in Japanese men. *Cancer Lett* 2001;164(1):33-40.

40. Hegele RA. Plasma lipoproteins: genetic influences and clinical implications. *Nat Rev Genet* 2009;10(2):109-121.

41. Teslovich TM, Musunuru K, Smith AV, et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* 2010;466(7307):707-713.

42. Meza R, Jeon J, Renehan AG, et al. Colorectal cancer incidence trends in the United States and United kingdom: evidence of right- to left-sided biological gradients with implications for screening. *Cancer Res* 2010;70(13):5419-5429.

43. Goff DC, Jr., Bertoni AG, Kramer H, et al. Dyslipidemia prevalence, treatment, and control in the Multi-Ethnic Study of Atherosclerosis (MESA): gender, ethnicity, and coronary artery calcium. *Circulation* 2006;113(5):647-656.

44. Cholesterol Treatment Trialists' Collaboration. Efficacy and safety of more intensive lowering of LDL cholesterol: a meta-analysis of data from 170 000 participants in 26 randomised trials. *Lancet* 2010;376(9753):1670-1681.

45. Poynter JN, Gruber SB, Higgins PD, et al. Statins and the risk of colorectal cancer. *N Engl J Med* 2005;352(21):2184-2192.

46. Wei JT, Mott LA, Baron JA, et al. Reported use of 3-hydroxy-3-methylglutaryl coenzyme A reductase inhibitors was not associated with reduced recurrence of colorectal adenomas. *Cancer Epidemiol Biomarkers Prev* 2005;14(4):1026-1027.

47. Siddiqui AA, Nazario H, Mahgoub A, et al. The long-term use of statins is associated with a decreased incidence of adenomatous colon polyps. *Digestion* 2009;79(1):17-22.

48. Bertagnolli MM, Hsu M, Hawk ET, et al. Statin use and colorectal adenoma risk: results from the adenoma prevention with celecoxib trial. *Cancer Prev Res (Phila)* 2010;3(5):588-596.

49. Burnett-Hartman AN, Passarelli MN, Adams SV, et al. Differences in epidemiologic risk factors for colorectal adenomas and serrated polyps by lesion severity and anatomical site. *Am J Epidemiol* 2013;177(7):625-637.

50. Morimoto LM, Newcomb PA, Ulrich CM, et al. Risk factors for hyperplastic and adenomatous polyps: evidence for malignant potential? *Cancer Epidemiol Biomarkers Prev* 2002;11(10 Pt 1):1012-1018.

51. Chia VM, Newcomb PA, Lampe JW, et al. Leptin concentrations, leptin receptor polymorphisms, and colorectal adenoma risk. *Cancer Epidemiol Biomarkers Prev* 2007;16(12):2697-2703.

52. Wernli KJ, Newcomb PA, Wang Y, et al. Body size, IGF and growth hormone polymorphisms, and colorectal adenomas and hyperplastic polyps. *Growth Horm IGF Res* 2010;20(4):305-309.

53. Burnett-Hartman AN, Newcomb PA, Mandelson MT, et al. Colorectal polyp type and the association with charred meat consumption, smoking, and microsomal epoxide hydrolase polymorphisms. *Nutr Cancer* 2011;63(4):583-592.

54. Burnett-Hartman AN, Newcomb PA, Mandelson MT, et al. No evidence for human papillomavirus in the etiology of colorectal polyps. *Cancer Epidemiol Biomarkers Prev* 2011;20(10):2288-2297.

55. Adams SV, Newcomb PA, Burnett-Hartman AN, et al. Circulating 25-hydroxyvitamin-D and risk of colorectal adenomas and hyperplastic polyps. *Nutr Cancer* 2011;63(3):319-326.

56. Burnett-Hartman AN, Newcomb PA, Hutter CM, et al. Variation in the association between colorectal cancer susceptibility loci and colorectal polyps by polyp type. *Am J Epidemiol* 2014:kwu114.

57. Saunders KW, Davis RL, Stergachis A. Group Health Cooperative of Puget Sound. In: Strom BL, ed. *Pharmacoepidemiology*. Chichester, United Kingdom: John Wiley & Sons, Ltd, 2000:247-262.

58. Winawer SJ, Zauber AG. The advanced adenoma as the primary target of screening. *Gastrointest Endosc Clin N Am* 2002;12(1):1-9, v.

59. Reddy BS, Wynder EL. Metabolic epidemiology of colon cancer. Fecal bile acids and neutral sterols in colon cancer patients and patients with adenomatous polyps. *Cancer* 1977;39(6):2533-2539.

60. McMichael AJ, Potter JD. Host Factors in Carcinogenesis - Certain Bile-Acid Metabolic Profiles That Selectively Increase the Risk of Proximal Colon Cancer. *J Natl Cancer Inst* 1985;75(2):185-191.

61. Walsh BW, Schiff I, Rosner B, et al. Effects of postmenopausal estrogen replacement on the concentrations and metabolism of plasma lipoproteins. *N Engl J Med* 1991;325(17):1196-1204.

62. Potter JD, Bostick RM, Grandits GA, et al. Hormone replacement therapy is associated with lower risk of adenomatous polyps of the large bowel: The Minnesota Cancer Prevention Research Unit Case-control Study. *Cancer Epidemiol Biomark Prev* 1996;5(10):779-784.

63. Hague A, Manning AM, Hanlon KA, et al. Sodium butyrate induces apoptosis in human colonic tumour cell lines in a p53-independent pathway: implications for the possible role of dietary fibre in the prevention of large-bowel cancer. *Int J Cancer* 1993;55(3):498-505.

64. Clausen MR, Bonnen H, Mortensen PB. Colonic fermentation of dietary fibre to short chain fatty acids in patients with adenomatous polyps and colonic cancer. *Gut* 1991;32(8):923-928.

65. Bonovas S, Filioussi K, Flordellis CS, et al. Statins and the risk of colorectal cancer: a meta-analysis of 18 studies involving more than 1.5 million patients. *J Clin Oncol* 2007;25(23):3462-3468.

66. Lochhead P, Chan AT. Statins and colorectal cancer. *Clin Gastroenterol Hepatol* 2013;11(2):109-118; quiz e113-104.

67. Siddiqui A, Nazario HE, Patel M, et al. Reduction in low-density lipoprotein cholesterol levels during statin therapy is associated with a reduced incidence of advanced colon polyps. *Am J Med Sci* 2009;338(5):378-381.

68. Bertagnolli MM, Hsu M, Hawk ET, et al. Statin use and colorectal adenoma risk: results from the adenoma prevention with celecoxib trial. *Cancer Prev Res (Phila)* 2010;3(5):588-596.

69. Broughton T, Sington J, Beales ILP. Statin use is associated with a reduced incidence of colorectal adenomatous polyps. *Int J Colorectal Dis* 2013;28(4):469-476.

70. Hozo SP, Djulbegovic B, Hozo I. Estimating the mean and variance from the median, range, and the size of a sample. *BMC Med Res Methodol* 2005;5:13.

71. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986;7(3):177-188.

72. Higgins JP, Thompson SG, Deeks JJ, et al. Measuring inconsistency in meta-analyses. *BMJ* 2003;327(7414):557-560.

73. Egger M, Davey Smith G, Schneider M, et al. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997;315(7109):629-634.

74. Berkey CS, Hoaglin DC, Mosteller F, et al. A random-effects regression model for meta-analysis. *Stat Med* 1995;14(4):395-411.

75. Heitman SJ, Ronksley PE, Hilsden RJ, et al. Prevalence of adenomas and colorectal cancer in average risk individuals: a systematic review and meta-analysis. *Clin Gastroenterol Hepatol* 2009;7(12):1272-1278.

76. Tabuchi M, Kitayama J, Nagawa H. Hypertriglyceridemia is positively correlated with the development of colorectal tubular adenoma in Japanese men. *World J Gastroenterol* 2006;12(8):1261-1264.

77. Law MR, Wald NJ, Rudnicka AR. Quantifying effect of statins on low density lipoprotein cholesterol, ischaemic heart disease, and stroke: systematic review and meta-analysis. *BMJ* 2003;326(7404):1423.

78. Kritchevsky SB, Wilcosky TC, Morris DL, et al. Changes in plasma lipid and lipoprotein cholesterol and weight prior to the diagnosis of cancer. *Cancer Res* 1991;51(12):3198-3203.

79. Berry EM, Zimmerman J, Peser M, et al. Dietary fat, adipose tissue composition, and the development of carcinoma of the colon. *J Natl Cancer Inst* 1986;77(1):93-97.

80. Mannes GA, Maier A, Thieme C, et al. Relation between the frequency of colorectal adenoma and the serum cholesterol level. *N Engl J Med* 1986;315(26):1634-1638.

81. Demers RY, Neale AV, Demers P, et al. Serum cholesterol and colorectal polyps. *J Clin Epidemiol* 1988;41(1):9-13.

82. Wang YY, Lin SY, Lai WA, et al. Association between adenomas of rectosigmoid colon and metabolic syndrome features in a Chinese population. *J Gastroenterol Hepatol* 2005;20(9):1410-1415.

83. Chung YW, Han DS, Park YK, et al. Association of obesity, serum glucose and lipids with the risk of advanced colorectal adenoma and cancer: a case-control study in Korea. *Dig Liver Dis* 2006;38(9):668-672.

84. Otani T, Iwasaki M, Ikeda S, et al. Serum triglycerides and colorectal adenoma in a case-control study among cancer screening examinees (Japan). *Cancer Causes Control* 2006;17(10):1245-1252.

85. Kuklina EV, Yoon PW, Keenan NL. Trends in high levels of low-density lipoprotein cholesterol in the United States, 1999-2006. *JAMA* 2009;302(19):2104-2110.

86. Cohen JD, Cziraky MJ, Cai Q, et al. 30-year trends in serum lipids among United States adults: results from the National Health and Nutrition Examination Surveys II, III, and 1999–2006. *Am J Cardiol* 2010;106(7):969-975.

87. Muldoon MF, Manuck SB, Matthews KA. Lowering cholesterol concentrations and mortality: a quantitative review of primary prevention trials. *BMJ* 1990;301(6747):309-314.

88. Patterson RE, Frank LL, Kristal AR, et al. A comprehensive examination of health conditions associated with obesity in older adults. *Am J Prev Med* 2004;27(5):385-390.

89. Wassertheil-Smoller S, Psaty B, Greenland P, et al. Association between cardiovascular outcomes and antihypertensive drug treatment in older women. *JAMA* 2004;292(23):2849-2859.

90. Howard BV, Kuller L, Langer R, et al. Risk of cardiovascular disease by hysterectomy status, with and without oophorectomy: the Women's Health Initiative Observational Study. *Circulation* 2005;111(12):1462-1470.

91. Beckman JA, Preis O, Ridker PM, et al. Comparison of usefulness of inflammatory markers in patients with versus without peripheral arterial disease in predicting adverse cardiovascular outcomes (myocardial infarction, stroke, and death). *Am J Cardiol* 2005;96(10):1374-1378.

92. Mainous AG, 3rd, Koopman RJ, Diaz VA, et al. A coronary heart disease risk score based on patient-reported information. *Am J Cardiol* 2007;99(9):1236-1241.

93. Hsia J, Margolis KL, Eaton CB, et al. Prehypertension and cardiovascular disease risk in the Women's Health Initiative. *Circulation* 2007;115(7):855-860.

94. Cook NR, Albert CM, Gaziano JM, et al. A randomized factorial trial of vitamins C and E and beta carotene in the secondary prevention of cardiovascular events in women: results from the Women's Antioxidant Cardiovascular Study. *Arch Intern Med* 2007;167(15):1610-1618.

95. Zhang C, Rexrode KM, van Dam RM, et al. Abdominal obesity and the risk of all-cause, cardiovascular, and cancer mortality: sixteen years of follow-up in US women. *Circulation* 2008;117(13):1658-1667.

96. Fortuny J, Sima C, Bayuga S, et al. Risk of endometrial cancer in relation to medical conditions and medication use. *Cancer Epidemiol Biomarkers Prev* 2009;18(5):1448-1456.

97. Pencina MJ, Navar-Boggan AM, D'Agostino RB, Sr., et al. Application of new cholesterol guidelines to a population-based sample. *N Engl J Med* 2014;370(15):1422-1431.

98. Stone NJ, Merz CNB, ScM F, et al. 2013 ACC/AHA Guideline on the Treatment of Blood Cholesterol to Reduce Atherosclerotic Cardiovascular Risk in Adults. *J Am Coll Cardiol* 2013 November 12 (Epub ahead of print).

99. Huang PY, Buring JE, Ridker PM, et al. Awareness, accuracy, and predictive validity of self-reported cholesterol in women. *J Gen Intern Med* 2007;22(5):606-613.

100. Kehoe R, Wu SY, Leske MC, et al. Comparing self-reported and physician-reported medical history. *Am J Epidemiol* 1994;139(8):813-818.

101. Gordon NF, Salmon RD, Franklin BA, et al. Effectiveness of therapeutic lifestyle changes in patients with hypertension, hyperlipidemia, and/or hyperglycemia. *Am J Cardiol* 2004;94(12):1558-1561.

102. Copeland KT, Checkoway H, McMichael AJ, et al. Bias due to misclassification in the estimation of relative risk. *Am J Epidemiol* 1977;105(5):488-495.

103. Flegal KM, Brownie C, Haas JD. The effects of exposure misclassification on estimates of relative risk. *Am J Epidemiol* 1986;123(4):736-751.

104. Jurek AM, Maldonado G, Greenland S, et al. Exposure-measurement error is frequently ignored when interpreting epidemiologic study results. *Eur J Epidemiol* 2006;21(12):871-876.

105. Bowlin SJ, Morrill BD, Nafziger AN, et al. Validity of cardiovascular disease risk factors assessed by telephone survey: the Behavioral Risk Factor Survey. *J Clin Epidemiol* 1993;46(6):561-571.

106. Nieto FJ, Alonso J, Chambless LE, et al. Population awareness and control of hypertension and hypercholesterolemia. The Atherosclerosis Risk in Communities study. *Arch Intern Med* 1995;155(7):677-684.

107. Murdoch M, Wilt TJ. Cholesterol awareness after case-finding: do patients really know their cholesterol numbers? *Am J Prev Med* 1997;13(4):284-289.

108. Natarajan S, Lipsitz SR, Nietert PJ. Self-report of high cholesterol: determinants of validity in U.S. adults. *Am J Prev Med* 2002;23(1):13-21.

109. Ford ES, Mokdad AH, Giles WH, et al. Serum total cholesterol concentrations and awareness, treatment, and control of hypercholesterolemia among US adults: findings from the National Health and Nutrition Examination Survey, 1999 to 2000. *Circulation* 2003;107(17):2185-2189.

110. He J, Gu D, Reynolds K, et al. Serum total and lipoprotein cholesterol levels and awareness, treatment, and control of hypercholesterolemia in China. *Circulation* 2004;110(4):405-411.

111. Muntner P, DeSalvo KB, Wildman RP, et al. Trends in the prevalence, awareness, treatment, and control of cardiovascular disease risk factors among noninstitutionalized patients with a history of myocardial infarction and stroke. *Am J Epidemiol* 2006;163(10):913-920.

112. Croyle RT, Loftus EF, Barger SD, et al. How well do people recall risk factor test results? Accuracy and bias among cholesterol screening participants. *Health Psychol* 2006;25(3):425-432.

113. Hyre AD, Muntner P, Menke A, et al. Trends in ATP-III-defined high blood cholesterol prevalence, awareness, treatment and control among US adults. *Ann Epidemiol* 2007;17(7):548-555.

114. Ford ES, Li CY, Pearson WS, et al. Trends in hypercholesterolemia, treatment and control among United States adults. *Int J Cardiol* 2010;140(2):226-235.

115. Egan BM, Li J, Qanungo S, et al. Blood pressure and cholesterol control in hypertensive hypercholesterolemic patients: national health and nutrition examination surveys 1988-2010. *Circulation* 2013;128(1):29-41.

116. Muntner P, Levitan EB, Brown TM, et al. Trends in the prevalence, awareness, treatment and control of high low density lipoprotein-cholesterol among United States adults from 1999–2000 through 2009–2010. *Am J Cardiol* 2013.

117. Colditz GA, Martin P, Stampfer MJ, et al. Validation of questionnaire information on risk factors and disease outcomes in a prospective cohort study of women. *Am J Epidemiol* 1986;123(5):894-900.

118. Newell S, Girgis A, Sanson-Fisher R, et al. Accuracy of patients' recall of Pap and cholesterol screening. *Am J Public Health* 2000;90(9):1431-1435.

119. Martin LM, Leff M, Calonge N, et al. Validation of self-reported chronic conditions and health services in a managed care population. *Am J Prev Med* 2000;18(3):215-218.

120. Merkin SS, Cavanaugh K, Longenecker JC, et al. Agreement of self-reported comorbid conditions with medical and physician reports varied by disease among end-stage renal disease patients. *J Clin Epidemiol* 2007;60(6):634-642.

121. Psaty BM, Lee M, Savage PJ, et al. Assessing the Use of Medications in the Elderly - Methods and Initial Experience in the Cardiovascular Health Study. *J Clin Epidemiol* 1992;45(6):683-692.

122. Lau HS, de Boer A, Beuning KS, et al. Validation of pharmacy records in drug exposure assessment. *J Clin Epidemiol* 1997;50(5):619-625.

123. Boudreau DM, Daling JR, Malone KE, et al. A validation study of patient interview data and pharmacy records for antihypertensive, statin, and antidepressant medication use among older women. *Am J Epidemiol* 2004;159(3):308-317.

124. Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. *Clin Chem* 1972;18(6):499-502.

125. Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults. Executive summary of the third report of The National Cholesterol Education Program (NCEP) Expert panel on detection, evaluation, and treatment of high blood cholesterol in adults (Adult Treatment Panel III). *JAMA* 2001;285(19):2486-2497.

126. Castelli WP. Epidemiology of triglycerides: a view from Framingham. *Am J Cardiol* 1992;70(19):3H-9H.

127. Goldberg JD. The effects of misclassification on the bias in the difference between two proportions and the relative odds in the fourfold table. *J Am Statist Assoc* 1975;70(351a):561-567.

128. Centers for Disease Control and Prevention. Vital signs: prevalence, treatment, and control of high levels of low-density lipoprotein cholesterol--United States, 1999-2002 and 2005-2008. *Morb Mortal Wkly Rep* 2011;60(4):109.

129. Psaty BM, Lee M, Savage PJ, et al. Assessing the use of medications in the elderly - methods and initial experience in the Cardiovascular Health Study. *J Clin Epidemiol* 1992;45(6):683-692.

130. Tirschwell DL, Smith NL, Heckbert SR, et al. Association of cholesterol with stroke risk varies in stroke subtypes and patient subgroups. *Neurology* 2004;63(10):1868-1875.

131. Doggen CJ, Smith NL, Lemaitre RN, et al. Serum lipid levels and the risk of venous thrombosis. *Arterioscler Thromb Vasc Biol* 2004;24(10):1970-1975.

132. Boudreau DM, Koehler E, Rulyak SJ, et al. Cardiovascular medication use and risk for colorectal cancer. *Cancer Epidemiol Biomarkers Prev* 2008;17(11):3076-3080.

133. Sidhu D, Naugler C. Fasting time and lipid levels in a community-based population: a cross-sectional study. *Arch Intern Med* 2012;172(22):1707-1710.

134. Davidson MH. Strategies to improve Adult Treatment Panel III guideline adherence and patient compliance. *Am J Cardiol* 2002;89(5A):8C-20C; discussion 20C-22C.

135. Mosca L, Linfante AH, Benjamin EJ, et al. National study of physician awareness and adherence to cardiovascular disease prevention guidelines. *Circulation* 2005;111(4):499-510.

136. Sisson SD, Rice TN, Hughes MT. Physician knowledge of national cholesterol guidelines before and after an interactive curriculum. *Am J Cardiol* 2007;99(9):1234-1235.

137. Mann D, Reynolds K, Smith D, et al. Trends in statin use and low-density lipoprotein cholesterol levels among US adults: Impact of the 2001 National Cholesterol Education Program guidelines. *Ann Pharmacother* 2008;42(9):1208-1215.

138. Sbarouni E, Flevari P, Kroupis C, et al. The effects of raloxifene and simvastatin on plasma lipids and endothelium. *Cardiovasc Drugs Ther* 2003;17(4):319-323.

139. McLaren DB, McKenzie M, Duncan G, et al. Watchful waiting or watchful progression?: Prostate specific antigen doubling times and clinical behavior in patients with early untreated prostate carcinoma. *Cancer* 1998;82(2):342-348.

140. Politser P. Reliability, decision rules, and the value of repeated tests. *Med Decis Making* 1982;2(1):47-69.

141. Marshall RJ. The predictive value of simple rules for combining two diagnostic tests. *Biometrics* 1989:1213-1222.

142. Thompson ML. Assessing the diagnostic accuracy of a sequence of tests. *Biostatistics* 2003;4(3):341-351.

143. Pepe MS. *The Statistical Evaluation of Medical Tests for Classification and Prediction*. New York, NY: Oxford University Press; 2003.

144. Albert PS. Estimating diagnostic accuracy of multiple binary tests with an imperfect reference standard. *Stat Med* 2009;28(5):780-797.

145. Bishop YM. Effects of collapsing multidimensional contingency tables. *Biometrics* 1971;27(3):545-562.

146. White E, Armstrong BK, Saracci R. *Principles of Exposure Measurement in Epidemiology*. 2nd ed. New York, NY: Oxford University Press; 2008.

147. Kim BC, Shin A, Hong CW, et al. Association of colorectal adenoma with components of metabolic syndrome. *Cancer Causes Control* 2012;23(5):727-735.

148. Yang MH, Rampal S, Sung J, et al. The association of serum lipids with colorectal adenomas. *Am J Gastroenterol* 2013;108(5):833-841.

149. Correa P, Strong JP, Johnson WD, et al. Atherosclerosis and polyps of the colon. Quantification of precursors of coronary heart disease and colon cancer. *J Chronic Dis* 1982;35(5):313-320.

150. Bernstein H, Bernstein C, Payne CM, et al. Bile acids as carcinogens in human gastrointestinal cancers. *Mutat Res* 2005;589(1):47-65.

151. Law MR, Thompson SG. Low serum cholesterol and the risk of cancer: an analysis of the published prospective studies. *Cancer Causes Control* 1991;2(4):253-261.

152. Cholesterol Treatment Trialists' Collaboration, Emberson, J. R., Kearney, P. M., et al. Lack of effect of lowering LDL cholesterol on cancer: meta-analysis of individual data from 175,000 people in 27 randomised trials of statin therapy. *PLoS One* 2012;7(1):e29849.

153. Glymour MM, Tchetgen EJ, Robins JM. Credible Mendelian randomization studies: approaches for evaluating the instrumental variable assumptions. *Am J Epidemiol* 2012;175(4):332-339.

154. Carroll MD, Kit BK, Lacher DA, et al. Trends in Lipids and Lipoproteins in US Adults, 1988-2010 Trends in Lipids and Lipoproteins in US Adults. *JAMA* 2012;308(15):1545-1554.

155. Heller DA, de Faire U, Pedersen NL, et al. Genetic and environmental influences on serum lipid levels in twins. *N Engl J Med* 1993;328(16):1150-1156.

156. Burgess S. Identifying the odds ratio estimated by a two-stage instrumental variable analysis with a logistic regression model. *Stat Med* 2013.

157. Palmer TM, Thompson JR, Tobin MD, et al. Adjusting for bias and unmeasured confounding in Mendelian randomization studies with binary responses. *Int J Epidemiol* 2008;37(5):1161-1168.

158. Burgess S, Thompson SG. Use of allele scores as instrumental variables for Mendelian randomization. *Int J Epidemiol* 2013;42(4):1134-1144.

159. Thompson SG, Higgins J. How should meta-regression analyses be undertaken and interpreted? *Stat Med* 2002;21(11):1559-1573.

160. Lawlor DA, Harbord RM, Sterne JA, et al. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Stat Med* 2008;27(8):1133-1163.

161. Tobin MD, Sheehan NA, Scurrah KJ, et al. Adjusting for treatment effects in studies of quantitative traits: antihypertensive therapy and systolic blood pressure. *Stat Med* 2005;24(19):2911-2935.

162. Katan MB. Apolipoprotein E isoforms, serum cholesterol, and cancer. *Lancet* 1986;1(8479):507-508.

163. Trompet S, Jukema JW, Katan MB, et al. Apolipoprotein e genotype, plasma cholesterol, and cancer: a Mendelian randomization study. *Am J Epidemiol* 2009;170(11):1415-1421.

164. Benn M, Tybjaerg-Hansen A, Stender S, et al. Low-density lipoprotein cholesterol and the risk of cancer: a mendelian randomization study. *J Natl Cancer Inst* 2011;103(6):508-519.

165. Solovieff N, Cotsapas C, Lee PH, et al. Pleiotropy in complex traits: challenges and strategies. *Nat Rev Genet* 2013;14(7):483-495.

166. Cai B, Small DS, Have TR. Two-stage instrumental variable methods for estimating the causal odds ratio: analysis of bias. *Stat Med* 2011;30(15):1809-1824.

167. Burgess S, Thompson SG. Bias in causal estimates from Mendelian randomization studies with weak instruments. *Stat Med* 2011;30(11):1312-1323.

168. Butler WJ, Ryan P, Roberts-Thomson IC. Metabolic genotypes and risk for colorectal cancer. *J Gastroenterol Hepatol* 2001;16(6):631-635.

169. Slattery ML, Sweeney C, Murtaugh M, et al. Associations between apoE genotype and colon and rectal cancer. *Carcinogenesis* 2005;26(8):1422-1429.

170. Brockton N, Little J, Sharp L, et al. N-acetyltransferase polymorphisms and colorectal cancer: a HuGE review. *Am J Epidemiol* 2000;151(9):846-861.

171. Lang NP, Butler MA, Massengill J, et al. Rapid metabolic phenotypes for acetyltransferase and cytochrome P4501A2 and putative exposure to food-borne heterocyclic amines increase the risk for colorectal cancer or polyps. *Cancer Epidemiol Biomarkers Prev* 1994;3(8):675-682.

172. Potter JD, Bigler J, Fosdick L, et al. Colorectal adenomatous and hyperplastic polyps: smoking and N-acetyltransferase 2 polymorphisms. *Cancer Epidemiol Biomarkers Prev* 1999;8(1):69-75.

173. Garcia-Closas M, Hein DW, Silverman D, et al. A single nucleotide polymorphism tags variation in the arylamine N-acetyltransferase 2 phenotype in populations of European background. *Pharmacogenet Genomics* 2011;21(4):231-236.

174. Rothman N, Garcia-Closas M, Chatterjee N, et al. A multi-stage genome-wide association study of bladder cancer identifies multiple susceptibility loci. *Nat Genet* 2010;42(11):978-984.

175. Chambers JC, Elliott P, Zabaneh D, et al. Common genetic variation near MC4R is associated with waist circumference and insulin resistance. *Nat Genet* 2008;40(6):716-718.

176. Speliotes EK, Willer CJ, Berndt SI, et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet* 2010;42(11):937-948.

177. Tenesa A, Campbell H, Theodoratou E, et al. Common genetic variants at the MC4R locus are associated with obesity, but not with dietary energy intake or colorectal cancer in the Scottish population. *Int J Obes (Lond)* 2009;33(2):284-288.

178. Okabayashi K, Ashrafian H, Hasegawa H, et al. Body Mass Index Category as a Risk Factor for Colorectal Adenomas: A Systematic Review and Meta-Analysis. *Am J Gastroenterol* 2012;107(8):1175-1185.

179. Broderick P, Chubb D, Johnson DC, et al. Common variation at 3p22.1 and 7p15.3 influences multiple myeloma risk. *Nat Genet* 2012;44(1):58-U83.

180. Crous-Bou M, Rennert G, Salazar R, et al. Genetic polymorphisms in fatty acid metabolism genes and colorectal cancer. *Mutagenesis* 2012;27(2):169-176.

181. Lipkin SM, Chao EC, Moreno V, et al. Genetic variation in 3-hydroxy-3-methylglutaryl CoA reductase modifies the chemopreventive activity of statins for colorectal cancer. *Cancer Prev Res (Phila)* 2010;3(5):597-603.

182. Vansteelandt S, Goetgeluk S, Lutz S, et al. On the adjustment for covariates in genetic association analysis: a novel, simple principle to infer direct causal effects. *Genet Epidemiol* 2009;33(5):394-405.

183. Wang J, Carvajal-Carmona LG, Chu JH, et al. Germline variants and advanced colorectal adenomas: adenoma prevention with celecoxib trial genome-wide association study. *Clin Cancer Res* 2013;19(23):6430-6437.

184. Edwards TL, Shrubsole MJ, Cai Q, et al. Genome-wide association study identifies possible genetic risk factors for colorectal adenomas. *Cancer Epidemiol Biomarkers Prev* 2013;22(7):1219-1226.

185. Peters U, Jiao S, Schumacher FR, et al. Identification of Genetic Susceptibility Loci for Colorectal Tumors in a Genome-Wide Meta-analysis. *Gastroenterology* 2013;144(4):799-807 e724.

186. Global Lipids Genetics Consortium. Discovery and refinement of loci associated with lipid levels. *Nat Genet* 2013;45(11):1274-1283.

187. Pierce BL, Ahsan H, Vanderweele TJ. Power and instrument strength requirements for Mendelian randomization studies using multiple genetic variants. *Int J Epidemiol* 2011;40(3):740-752.

188. Eichler EE, Flint J, Gibson G, et al. Missing heritability and strategies for finding the underlying causes of complex disease. *Nat Rev Genet* 2010;11(6):446-450.

189. Miller JI, Ahmann FR, Drach GW, et al. The clinical usefulness of serum prostate specific antigen after hormonal therapy of metastatic prostate cancer. *J Urol* 1992;147(3 Pt 2):956-961.

190. Taylor F, Huffman MD, Macedo AF, et al. Statins for the primary prevention of cardiovascular disease. *Cochrane Database Syst Rev* 2013;1(1):CD004816.

191. Demierre MF, Higgins PD, Gruber SB, et al. Statins and cancer prevention. *Nat Rev Cancer* 2005;5(12):930-942.

192. Welch HG, Albertsen PC, Nease RF, et al. Estimating treatment benefits for the elderly: the effect of competing risks. *Ann Intern Med* 1996;124(6):577-584.

193. Wright AF, Carothers AD, Pirastu M. Population choice in mapping genes for complex diseases. *Nat Genet* 1999;23(4):397-404.

194. Vickers AJ, Till C, Tangen CM, et al. An empirical evaluation of guidelines on prostate-specific antigen velocity in prostate cancer detection. *J Natl Cancer Inst* 2011.

195. Borofsky MS, Makarov DV. Prostate cancer: PSA velocity in prostate cancer screening--the debate continues. *Nat Rev Urol* 2011;8(8):413-414.

196. Kontush A, Chapman MJ. Functionally defective high-density lipoprotein: a new therapeutic target at the crossroads of dyslipidemia, inflammation, and atherosclerosis. *Pharmacol Rev* 2006;58(3):342-374.

197. Rosenson RS, Brewer HB, Jr., Chapman MJ, et al. HDL measures, particle heterogeneity, proposed nomenclature, and relation to atherosclerotic cardiovascular events. *Clin Chem* 2011;57(3):392-410.

198. Heinecke J. HDL and cardiovascular-disease risk--time for a new approach? *N Engl J Med* 2011;364(2):170-171.

199. Bhattacharyya T, Nicholls SJ, Topol EJ, et al. Relationship of paraoxonase 1 (PON1) gene polymorphisms and functional activity with systemic oxidative stress and cardiovascular risk. *JAMA* 2008;299(11):1265-1276.

200. Willer CJ, Sanna S, Jackson AU, et al. Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat Genet* 2008;40(2):161-169.

201. Voight BF, Peloso GM, Orho-Melander M, et al. Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study. *Lancet* 2012;380(9841):572-580.

202. Khera AV, Cuchel M, de la Llera-Moya M, et al. Cholesterol efflux capacity, high-density lipoprotein function, and atherosclerosis. *N Engl J Med* 2011;364(2):127-135.

203. Francis GA, Knopp RH, Oram JF. Defective removal of cellular cholesterol and phospholipids by apolipoprotein AI in Tangier Disease. *J Clin Invest* 1995;96(1):78.

204. Zakai NA, Katz R, Jenny NS, et al. Inflammation and hemostasis biomarkers and cardiovascular risk in the elderly: the Cardiovascular Health Study. *J Thromb Haemost* 2007;5(6):1128-1135.

205. Nelson ER, Wardell SE, Jasper JS, et al. 27-Hydroxycholesterol links hypercholesterolemia and breast cancer pathophysiology. *Science* 2013;342(6162):1094-1098.

206. Gutfeld O, Prus D, Ackerman Z, et al. Expression of serum amyloid A, in normal, dysplastic, and neoplastic human colonic mucosa: implication for a role in colonic tumorigenesis. *J Histochem Cytochem* 2006;54(1):63-73.

207. Benjamini Y, Hochberg Y. Controlling the false discovery rate - a practical and powerful approach to multiple testing. *J R Stat Soc B* 1995;57(1):289-300.

208. Cook NR. Imputation strategies for blood pressure data nonignorably missing due to medication use. *Clinical Trials* 2006;3(5):411-420.

209. Levy D, Ehret GB, Rice K, et al. Genome-wide association study of blood pressure and hypertension. *Nat Genet* 2009;41(6):677-687.

210. Little J, Higgins J, Ioannidis J, et al. STrengthening the REporting of Genetic Association studies (STREGA)–an extension of the STROBE statement. *Eur J Clin Invest* 2009;39(4):247-266.

211. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med* 2007;147(8):573-577.

212. Masca N, Sheehan NA, Tobin MD. Pharmacogenetic interactions and their potential effects on genetic analyses of blood pressure. *Stat Med* 2011;30(7):769-783.

Vita

Michael Nicholas Passarelli is originally from Syracuse, New York. He received a Bachelor of Science degree in mathematics from Bucknell University in 2003, a Master of Science degree in medical statistics from University of Rochester in 2005, a Master of Public Health degree in biostatistics from University of Michigan in 2009, and a Doctor of Philosophy degree in epidemiology from University of Washington in 2014. He will continue training in genetic and molecular epidemiology of cancer as a post-doctoral researcher at University of California, San Francisco.