Clinical Chemistry 57:3
502–510 (2011)

Use of Genotype Frequencies in Medicated Groups to Investigate Prescribing Practice: *APOE* and Statins as a Proof of Principle

Neil M. Davies,[1] Frank Windmeijer,[2] Richard M. Martin,[1] Mohammad R. Abdollahi,[1] George Davey Smith,[1] Debbie A. Lawlor,[1] Shah Ebrahim,[3] and Ian N.M. Day[1*]

BACKGROUND: If treatments are used to modify a trait, then patients with high-risk genotypes for the trait should be found at higher frequency in treatment groups than in the general population. The frequency ratio of high- to low-risk genotypes treated should reflect the mean threshold above which the treatment is given in the population. As an example, we hypothesized that because *APOE* (apolipoprotein E) alleles affect the LDL cholesterol (LDLc) concentration, *APOE* genotype frequencies in statin takers should act as a proxy for the prevailing treatment threshold of LDLc.

METHODS: We used LDLc, statin usage, and *APOE* genotype data from the British Women's Heart and Health Study (n = 2289; age, 60–79 years) and calculated the genotype ratio treatment index (GRTI) by dividing the proportion of $\varepsilon3/\varepsilon2$ or $\varepsilon3/\varepsilon4$ participants prescribed a statin by the proportion of $\varepsilon3/\varepsilon3$ participants prescribed a statin, both overall and according to socioeconomic class, geographic region, and coronary heart disease (CHD) status. Genotype-specific LDLc distributions were used to calculate the mean LDLc treatment threshold.

RESULTS: For genotype $\varepsilon3/\varepsilon2$, the GRTI was 0.52 (95% CI, 0.30–0.87) for statin takers overall, 0.22 (95% CI, 0.00–0.56) for those without CHD, and 0.69 (95% CI, 0.31–1.18) for those with CHD. The GRTIs for those without and with CHD backcalculate to LDLc thresholds of 5.65 mmol/L (95% CI, 5.50–5.82 mmol/L) and 4.39 mmol/L (95% CI, 4.21–4.59 mmol/L), respectively. Scotland and North England showed dissimilar GRTIs, which backcalculated to LDLc thresholds of 5.06 mmol/L (95% CI, 4.83–5.28 mmol/L) and 5.44 mmol/L (95% CI, 5.19–5.69 mmol/L), respectively, for all women.

CONCLUSIONS: The findings illustrate how genotype frequencies can be a proxy for treatment thresholds used in clinical practice. Genome-wide studies have identified >500 disease-relevant polymorphisms. GRTIs from cost-efficient genotyping, in combination with phenotypic data, may have wide potential in health services research.

A substantial proportion of disease is attributable to inequalities in healthcare between countries, within countries and regions, and between subgroups of people. In contrast, the distributions of common polymorphisms in the general population appear to be independent of individuals' backgrounds *(1, 2)*; however, if a genotype raises the level of a risk trait or disease and treatment is used to modify the trait, then patients with that genotype should be found at a higher frequency in the treatment group than in the general population. This difference should reflect the mean threshold of the trait above which the intervention is being used in a specific population, as, for example, members of a social class, residents of a specific geographic area, or patients admitted to a particular hospital. Furthermore, for common polymorphisms with minor effects on common diseases, the genotype usually remains unknown during a patient's clinical presentation and subsequent management. Thus, analogous to double-blinded clinical trials *(3)*, which involve the use of participant randomization, allocation concealment, and blinding of investigators to estimate the effect of a trialed intervention, a genotype frequency–based assessment of prescribing practice has the advantage of using genotype as the concealed variable, thus reducing confounding and other biases. Given the growing ease

[1] MRC Centre (CAiTE) and Bristol Genetic Epidemiology Laboratories, School of Social and Community Medicine, University of Bristol, Bristol, UK; [2] Centre for Market and Public Organisation, University of Bristol, Bristol, UK; [3] Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, UK.
* Address correspondence to this author at: MRC Centre (CAiTE) and Bristol Genetic Epidemiology Laboratories, School of Social and Community Medicine, University of Bristol, Oakfield House, Oakfield Grove, Bristol BS8 2BN, UK. E-mail ian.day@bristol.ac.uk.
Received October 29, 2010; accepted November 4, 2010.
Previously published online at DOI: 10.1373/clinchem.2010.156356

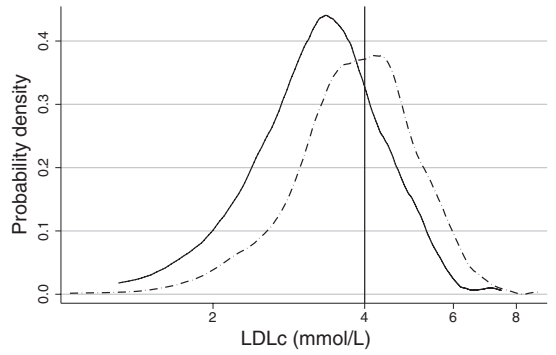

Fig. 1. Distribution of LDLc values in the BWHHS cohort for 2 *APOE* genotypes.

LDLc is plotted on a logarithmic scale. Note that the proportion of individuals of ε3/ε2 genotype above any given LDLc concentration threshold value (e.g., 4 mmol/L) will be less than that for individuals of ε3/ε3 genotype. The ratio of these proportions (i.e., the GRTI) will be low at a high LDLc threshold but will increase gradually to 1 as the LDLc threshold is decreased toward 0 mmol/L. Solid line, *APOE* genotype ε3/ε2; broken line, *APOE* genotype ε3/ε3. To convert LDLc values in millimoles per liter to milligrams per deciliter, multiply by 38.61.

of genome-wide typing and the numerous polymorphisms that might act as proxies in this way, such an approach has potential to be an important generic method for investigating healthcare delivery.

For example, common polymorphisms that affect LDL cholesterol (LDLc)[4] concentration *(4)* should affect the likelihood of being prescribed a statin, although such genotypes are not useful when making treatment decisions. *APOE* (apolipoprotein E) genotype represents a classic example. The group with genotype ε3/ε2 displays a lower mean LDLc (and total cholesterol) concentration than the group with genotype ε3/ε3. For a given LDLc threshold (which might be a treatment threshold or part of a multiple–risk factor decision index), a smaller proportion of the individuals in the ε3/ε2 genotype group will fall above the threshold than in the ε3/ε3 genotype group. These proportions will depend on the actual threshold and should be uniquely identifying of that threshold for a homogeneous population (see Fig. 1). Thus, the genotype proportions alone should act as a proxy for the mean threshold for treatment in the population [i.e., a genotype ratio treatment index (GRTI)]. Aspects of population heterogeneity are considered in the discussion. The purpose of this report is to illustrate that the GRTI may be a new tool that could be added to those used to index treatment decisions or treatment events at the population level. For example, it could enable comparisons of healthcare delivery to different subgroups (e.g., prescribing statins in the case of *APOE*), such as individuals in different social classes, residents of different geographic areas, or individuals with different disease risks. Furthermore, given knowledge of the underlying distributions of LDLc (or total cholesterol) concentration for (untreated) genotype groups, it should then be possible to estimate the mean prescribing threshold for the original sample.

We used data from a cohort study in a proof-of-principle analysis of the following: LDLc concentration as the clinical scenario; statin prescribing as the therapeutic decision; *APOE* genotype frequency ratios as the tool to assess the mean treatment threshold used in a population; and geographical, social, and disease-status groups as exemplary strata within which to apply the proposed GRTI.

Methods

DATA USED FOR PROOF-OF-PRINCIPLE ANALYSIS

We used data from a population-based cohort, the British Women's Heart and Health Study (BWHHS), which include data on LDLc concentration, *APOE* genotype, and statin prescriptions, as well as a variety of other clinical, geographic, and socioeconomic information. All participants included in the analyses provided written consent, and local and central ethics committees approved the study. Between 1999 and 2001, 4286 women (age, 60–79 years) from 23 British towns were randomly selected from the lists of general practitioners and then interviewed and examined. The women completed medical questionnaires, a nurse interview (including a review of medications), and a physical examination, during which blood samples were taken (after minimum 6 h of fasting) and their medical records were reviewed. The total cholesterol, HDL cholesterol (HDLc), and triglyceride concentrations in frozen serum samples (maximum time frozen, 6 weeks) were measured with a Hitachi 757 analyzer (Roche Diagnostics) and standard reagent sets. The LDLc concentration was estimated from the Friedewald equation [LDLc = total cholesterol − HDLc − 0.45(TG), where TG is the triglyceride concentration and concentrations are expressed in millimoles per liter] *(5)*. For total cholesterol, the within-batch CV was 0.96% at 5.11 mmol/L and 0.81% at 7.21 mmol/L (both estimates based on 20 replicates), and the between-batch CV was 1.3% at 5.24 mmol/L and 7.21 mmol/L

[4] Nonstandard abbreviations: LDLc, LDL cholesterol; GRTI, genotype ratio treatment index; BWHHS, British Women's Heart and Health Study; HDLc, HDL cholesterol; CHD, coronary heart disease.

(based on 13 replicates). We tested the sensitivity of the results to imprecision in the Friedewald equation by restricting the analyses to participants with triglyceride concentrations <2.5 mmol/L. *APOE* genotype was determined for 3271 individuals *(4)*. There are 3 common alleles at the *APOE* locus—ε2, ε3, and ε4. Analyses were restricted to the common *APOE* genotypes ε3/ε3, ε3/ε2, and ε3/ε4. Coronary heart disease (CHD) cases included any self-reported or doctor diagnosis of angina or myocardial infarction, or any evidence in the medical records of either angina or myocardial infarction, which was determined at the same time as statin-taker status. Adult social class was derived from the longest held position for single women and from that of the participant's husband for married women. Social class was split into groups according to the register general's classification *(6)*. The geographic analysis used a division into northern England, the Midlands, southern England, and Scotland, as previously described *(7)*. *APOE* genotype and allele frequency data were extracted from published large-scale studies in the UK *(8–13)* and, for illustrative comparison, from a Taiwanese study *(8)* and a Finnish study *(9)*.

GENOTYPE RATIO TREATMENT INDICES

The population frequencies of ε3/ε2, ε3/ε3, and ε3/ε4 genotypes were estimated from BWHHS data. These frequencies were compared with the population frequencies reported from other large-scale population-based studies in the UK. Frequencies of ε3/ε2, ε3/ε3, and ε3/ε4 genotypes were then determined in BWHHS subgroups (e.g., division by social class, presence or absence of CHD, geographic area), and these frequencies were compared with those for the overall BWHHS population. The frequencies of the same genotypes in statin takers were determined for the overall BWHHS population and for the subgroups of social class, geographic region, and CHD status. The GRTIs for these subgroups (ε3/ε2 relative to ε3/ε3 and ε3/ε4 relative to ε3/ε3) were calculated by dividing the proportion of ε3/ε2 (or ε3/ε4) participants prescribed a statin by the proportion of ε3/ε3 participants prescribed a statin.

CONVERTING GRTI BACK TO AN ESTIMATED MEAN LDLc TREATMENT THRESHOLD

Because LDLc distributions may vary by the stratifying variables, we used the LDLc data of population subgroups and estimated the treatment thresholds used in the subpopulations. Estimated LDLc data for BWHHS were plotted by genotype for *APOE* groups ε3/ε2, ε3/ε3, and ε3/ε4. Because of the effects of statin prescribing on a (small) subset of the population in the upper half of the distributions, we used Tobit regression analysis for censored data and treated the LDLc concentration as a log-normal variable to generate the distributions (see the Appendix in the Data Supplement that accompanies the online version of this article at http://www.clinchem.org/content/vol57/issue3). From these distributions, we calculated expected GRTIs (both for the ratio of ε3/ε2 to ε3/ε3 and the ratio of ε3/ε4 to ε3/ε3) across the entire range of possible LDLc threshold values. We then converted the observed GRTI back to an LDLc value, which represented the mean concentration threshold above which treatment was started in that group. For these conversions, we used the LDLc distribution by *APOE* genotype for the specific population subsample in which the threshold was being estimated. A standard likelihood function was used to combine pairs of GRTIs (for the ratio of ε3/ε2 to ε3/ε3 and the ratio of ε3/ε4 to ε3/ε3) for a given subgroup. The details of this analysis are presented in the Appendix in the online Data Supplement. We conducted a number of sensitivity analyses, including calculating the threshold with the HDLc/LDLc ratio (see the Appendix in the online Data Supplement). Statistical analysis was performed with Stata software (version 10.1; StataCorp).

Results

SAMPLE SELECTION AND POSSIBLE CONFOUNDERS

The 4286 participants represented a 60% response rate of those eligible *(9)*. The prevalence of CHD was similar in the responders and nonresponders, but for some conditions (e.g., diabetes and stroke) the responders were healthier than the nonresponders. Genotype data were available for 3271 participants, of which 862 women had some missing values for other covariates. Our analysis is confined to the 2289 women with relevant complete data. This sample of women had a mean body mass index that was 0.6 kg/m^2 lower than for the portion of the cohort that was excluded because of missing data ($P < 0.01$) and were 8% less likely to have any evidence of CHD ($P < 0.001$). There was no evidence that the included and excluded women differed with respect to LDLc concentration. The frequencies of the 6 *APOE* genotypes conformed closely to the Hardy–Weinberg equilibrium *(5)*. *APOE* genotype and allele frequencies were not significantly different from those of other published large UK studies (Table 1). By contrast, the *APOE* genotype and allele frequencies differed from those of a large sample of individuals of Han Chinese ancestry and subtly differed from those of a large Finnish sample. There was no association in the BWHHS between *APOE* genotype and socioeconomic class, geographic region of residence, age, or anthropometric measurements.

Table 1. *APOE* allele frequencies in cohort studies.

Study	ε2 Allele	ε3 Allele	ε4 Allele	Alleles of participants, n × 2	Reference
BWHHS (UK multicenter)	0.084	0.776	0.140	6542	Abdollahi et al. *(4)*
UK multicenter (NPHSII)[a] reference	0.078	0.771	0.151	4516	Humphries et al. *(13)*
Wessex, UK (SAS)	0.073	0.793	0.134	2310	Ye et al. *(14)*
PROSPER (Scotland, Ireland, the Netherlands)	0.074	0.792	0.135	11 088	Packard et al. *(15)*
Whitehall, male admin	0.0784	0.769	0.153	3966	Zhao et al. *(16)*
Whitehall, male prof/exec	0.0807	0.772	0.148	3060	Zhao et al. *(16)*
Whitehall, female admin	0.0712	0.780	0.149	1292	Zhao et al. *(16)*
Whitehall, female prof/exec	0.0717	0.785	0.143	2676	Zhao et al. *(16)*
Go-DARTS (Scotland)	0.0828	0.775	0.142	4902	Donnelly et al. *(12)*
Scotland reference (Grampian region)	0.08	0.77	0.15	400	Cumming and Robertson *(17)*
ALSPAC (Avon, UK)	0.088	0.761	0.151	11 990	Taylor et al. *(18)*
Finland reference	0.064	0.76	0.175	1880	Haddy et al. *(9)*
Han Chinese reference (Taiwan)	0.118	0.841	0.081	4652	Liu et al. *(8)*

[a] NPHSII, The second Northwick Park Heart Study; SAS, Southampton Atherosclerosis Study; PROSPER, Prospective Study of Pravastatin in the Elderly at Risk; admin, administration (secretarial) staff; prof, professional staff; exec, executive staff; Go-DARTS, Genetics of Diabetes Audit and Research Tayside Study; ALSPAC, Avon Longitudinal Study of Parents and Children.

APOE GENOTYPE, LDLc DISTRIBUTIONS, AND THE LIKELIHOOD OF BEING PRESCRIBED A STATIN

Fig. 1 shows the distributions of LDLc concentrations for genotypes ε3/ε3 and ε3/ε2 in the BWHHS women who were not taking statins. Relative to the ε3/ε3 curve, the ε3/ε2 curve is shifted to the left—toward lower LDLc concentrations. Thus, for any particular LDLc concentration (e.g., x set at 4 mmol/L in Figs. 1 and 2), the proportion of ε3/ε2 individuals with a value $>x$ is less than the proportion of ε3/ε3 individuals with a value $>x$. This is exactly what we would expect to see in the context of statin treatment. If all individuals displaying an LDLc concentration $>x$ were prescribed a statin, then the proportion of ε3/ε2 individuals receiving a statin will be less than the proportion of ε3/ε3 individuals receiving a statin. Furthermore, there will be a unique relationship between x and this ratio. Fig. 2 shows this relationship of GRTI to x. If an x value of 4.0 mmol/L were the mean prescribing threshold for a group receiving statins, the GRTI (i.e., the ratio of ε3/ε2 to ε3/ε3) would be 0.56.

GRTIs FOR STATINS

Table 2 presents the statin GRTIs for the ε3/ε2 and ε3/ε4. The ratio of the proportion of ε3/ε2 participants prescribed statins to the proportion of ε3/ε3 participants prescribed statins was 0.52 (95% CI, 0.30–0.87) for the full sample. The corresponding ratio for ε3/ε4 to ε3/ε3 was 1.46 (95% CI, 1.00–2.02). That is, ε3/ε4 individuals had an almost 50% greater likelihood of statin prescription than ε3/ε3 individuals, and ε3/ε2 individuals were approximately 50% as likely as ε3/ε3 individuals to be prescribed a statin. The GRTIs for the population subgroups are shown in Table 3, which includes the ratios for subgroups with or without a cor-

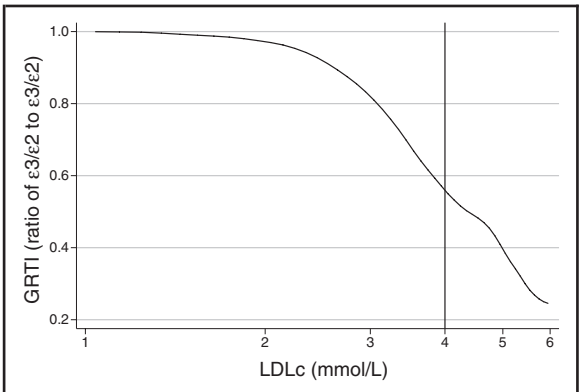

Fig. 2. Relationship between the LDLc concentration and the GRTI for ε3/ε2 individuals who are statin takers (from BWHHS cohort data).

A GRTI value of approximately 0.56 would read back to an estimated mean prescribing LDLc threshold of 4 mmol/L. The minor inflection in the curve likely reflects sparse data, considering the number of statin takers in this study. To convert LDLc values in millimoles per liter to milligrams per deciliter, multiply by 38.61.

Table 2. Statin GRTI values for the ratios of *APOE* genotype ε3/ε2 to ε3/ε3 and of *APOE* genotype ε3/ε4 to ε3/ε3.[a]

	Participants, n		GRTI, ratio (95% CI)	
Sample	Total	On statins	ε3/ε4 to ε3/ε3	ε3/ε2 to ε3/ε3
Full sample	2278	164	1.46 (1.00–2.02)	0.52 (0.30–0.87)
Without CHD at baseline	1971	72	1.41 (0.73–2.45)	0.22 (0.00–0.56)
With CHD at baseline	307	92	1.67 (1.20–2.19)	0.69 (0.31–1.18)
High social class	743	54	1.37 (0.70–2.33)	0.31 (0.00–0.75)
Low social class	1535	110	1.51 (0.97–2.18)	0.62 (0.30–1.00)
South England	726	52	1.75 (0.83–3.13)	0.66 (0.00–1.40)
North England	899	60	1.10 (0.54–2.04)	0.40 (0.00–1.12)
Scotland	305	27	2.68 (0.73–5.91)	0.81 (0.00–2.26)
Midlands	348	25	1.32 (0.00–2.85)	0.60 (0.00–1.23)
Age >70 years	901	71	2.02 (1.17–3.17)	0.76 (0.30–1.47)
Age <70 years	1377	93	1.17 (0.68–1.84)	0.39 (0.13–0.76)
BMI[b] >25 kg/m^2	1508	119	1.31 (0.83–1.88)	0.58 (0.24–1.02)
BMI <25 kg/m^2	770	45	1.98 (0.99–3.31)	0.38 (0.00–0.95)
Diastolic BP >80 mmHg	1013	64	1.70 (1.04–2.65)	0.31 (0.00–0.73)
Diastolic BP <80 mmHg	1265	100	1.34 (0.80–2.01)	0.62 (0.28–1.07)
Ever smoker	953	82	1.39 (0.82–2.15)	0.81 (0.38–1.32)
Never smoker	1325	82	1.56 (1.00–2.34)	0.28 (0.00–0.65)

[a] CIs obtained from 1000 bootstrap repetitions. A 0.00 value for the 2.5% confidence limit occurs when there were no ε3/ε2 prescribed statins in at least 2.5% of the bootstrapped repetitions.
[b] BMI, body mass index; BP, blood pressure.

onary event and subgroups by social class and geographic region. For example, the ε3/ε2 subgroup with CHD showed a statin GRTI of 0.69 (95% CI, 0.31–1.18), in contrast with the ε3/ε2 subgroup without CHD, which had a GRTI of 0.22 (95% CI, 0.00–0.56). In sensitivity analyses restricted to 1753 participants with triglyceride concentrations <2.5 mmol/L, the GRTIs did not differ substantively (see Table 1 in the online Data Supplement).

ESTIMATES OF THE LDLc MEAN PRESCRIBING THRESHOLD

Table 3 shows maximum likelihood estimates of the threshold; the estimate of the threshold for the full sample is 5.30 mmol/L (95% CI, 5.18–5.43 mmol/L). The estimated mean thresholds for social class and presence or absence of CHD are also shown. The mean threshold for receiving statins, which is calculated from the values for genotypic status entered into the likelihood function, is 5.65 mmol/L (95% CI, 5.50–5.82 mmol/L) for those without CHD, compared with 4.39 mmol/L (95% CI, 4.21–4.59 mmol/L) for those with CHD. The thresholds did not differ by social class, but the estimated mean thresholds for northern England and Scotland showed a trend toward a significant difference [5.44 mmol/L (95% CI, 5.19–5.69 mmol/L) and 5.06 mmol/L (95% CI, 4.83–5.28 mmol/L), respectively]. The thresholds decreased marginally when we restricted the sample to participants with triglyceride concentrations <2.5 mmol/L. This restriction changed the estimated LDLc thresholds by approximately 0.1 mmol/L (see Table 2 in the online Data Supplement).

Discussion

In this proof-of-principle study, we demonstrated the use of genotype frequencies in medicated groups to make inferences about the threshold at which prescribing decisions are made in clinical practice. This approach is useful because it offers a novel way to track healthcare delivery. It may be independent of specific measurements, such as cholesterol concentration, for which genotype–phenotype distribution data specific for a (homogeneous) population are available. If such distributions were unknown, then data on genotype and phenotype would be required to estimate a GRTI. The GRTI would identify heterogeneities in treatments at the population level. Further investigation might reveal these heterogeneities to represent appropriate or

Table 3. Estimates of the mean prescribing thresholds in LDLc.[a]

Sample	Participants, n		LDLc threshold, mmol/L	
	Total	On statins	Mean	95% CI
Full sample	2278	164	5.30	5.18–5.43
Without CHD at baseline	1971	72	5.65	5.50–5.82
With CHD at baseline	307	92	4.39	4.21–4.59
High social class	743	54	5.26	5.03–5.51
Low social class	1535	110	5.32	5.19–5.45
South England	726	52	5.23	5.02–5.46
North England	899	60	5.44	5.19–5.69
Scotland	305	27	5.06	4.83–5.28
Midlands	348	25	5.22	5.03–5.38
Age >70 years	901	71	5.37	5.15–5.60
Age <70 years	1377	93	5.25	5.13–5.39
BMI[b] >25 kg/m^2	1508	119	5.26	5.13–5.41
BMI <25 kg/m^2	770	45	5.39	5.21–5.59
Diastolic BP >80 mmHg	1013	64	5.45	5.27–5.68
Diastolic BP <80 mmHg	1265	100	5.19	5.05–5.33
Ever smoker	953	82	5.10	4.95–5.26
Never smoker	1325	82	5.47	5.32–5.65

[a] Mean threshold estimated with the likelihood function and according to the process described in Appendix I in the online Data Supplement. SEs were estimated by nonparametric bootstrapping with 1000 repetitions. Thresholds were estimated for $\epsilon3/\epsilon2$, $\epsilon3/\epsilon3$, and $\epsilon3/\epsilon4$ genotypes. To convert LDLc concentrations in millimoles per liter to milligrams per deciliter, multiply values by 38.61.
[b] BMI, body mass index; BP, blood pressure.

inappropriate inequality of treatment. Given the availability of population-based genome-wide data, numerous aspects of healthcare delivery might be investigated with GRTIs.

Fig. 3 illustrates a possible process for the use of GRTIs; however, important stratifying variables, such as socioeconomic category, can affect phenotype distributions independently of the genotype. Therefore, to use and interpret GRTIs by backcalculation to a threshold requires a population stratum–specific distribution, but if a group is homogeneous, the GRTI alone can enable comparison of treatment rates in subgroups. Because the genotype is generally unknown and because the GRTI offers a process similar to randomization in a drug trial, a GRTI assessment of healthcare delivery could be made without the usual problems of confounding and selection bias. The genotype will generally remain unknown because a genotype such as the *APOE* genotype is generally not useful information and is not clinically recognizable during initial clinical assessment or prescribing decisions.

Currently, there are large increments in genotyping throughput *(10)* and consequently increasing knowledge of the effects of gene polymorphisms *(2)*. The GRTI could have many uses in the research and auditing of prescribing patterns and other treatments. In the future, a once-in-a-lifetime genome typing process for a population would make GRTIs very cost-effective; however, either genotypes with a substantial effect or combinations of genotypes for different loci that have a substantial effect in aggregate are required. Additionally, if the results from different loci with unrelated pleiotropic effects all concur, then such results are probably due to common main effects, rather than to pleiotropy. Sufficient scaled sampling (ultimately as a national program) would ensure the statistical power of the approach.

We have shown that groups of individuals in a population sample with different *APOE* genotypes have different likelihoods of receiving statin therapy. Genotype frequencies alone can act as an index for treatment groups relative to population genotype frequencies, with the GRTIs reflecting the mean statin prescribing threshold in the subgroup of interest. We used known LDLc distributions by genotype for the population to estimate the GRTIs for subgroups and estimated the mean LDLc threshold used for prescribing in subgroups. Our proof-of-principle study used

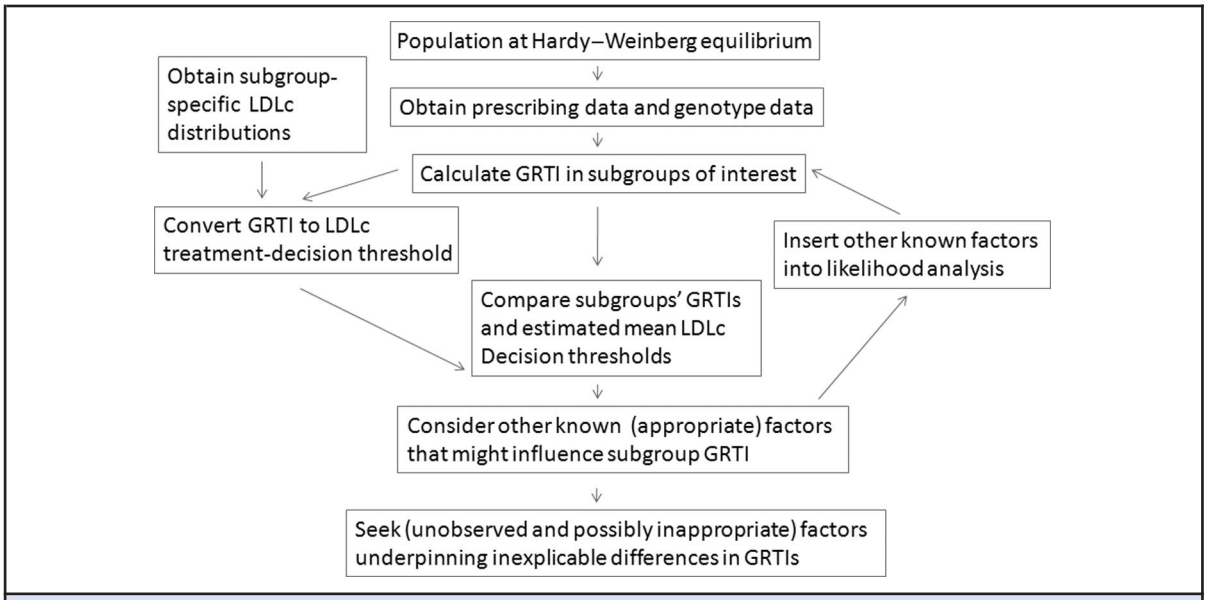

Fig. 3. Schematic illustrating the possible process of deployment of the GRTI.

data from a cohort study; however, in the future, the GRTIs would use population-level data to investigate many different aspects of healthcare delivery.

We found that a group with CHD and a subgroup without CHD had different GRTIs. We also showed that the genotypic likelihoods of receiving a statin could be used to backcalculate (parsimoniously with multiple genotypes in a likelihood function) the mean LDLc threshold above which statins are prescribed. The pattern of the GRTIs (with $\varepsilon3/\varepsilon2$ and $\varepsilon3/\varepsilon4$ indices responding in opposite directions to a change in threshold) showed that more-aggressive prescribing occurred in individuals with CHD, compared with the primary-prevention subgroup. Backcalculating via the likelihood function to an LDLc threshold revealed distinct CIs. The differences between social class subgroups were small. There was evidence (Table 3) that the Scotland and northern England geographic groups differed in prescribing, a difference that might reflect the influence of other risk factors on prescribing, variation in healthcare delivery, or different clinical guidance on statins in Scotland.

The key point is that the GRTIs highlight differences worthy of further investigation. The smallest detectable difference between northern England and Scotland in prescribing threshold with 85% power and a 95% level of confidence is 0.51 mmol/L. A national GRTI framework would allow higher-powered comparisons and would use more genetic variants. For example, at least 95 genotypes influence LDLc concentration *(17)*. For genotypes representing other traits or diseases, the power will depend on the sizes of the genotypic effects. Ultimately, genotypes such as those for *APOE* could be used at different points in the healthcare-delivery process (e.g., diagnosis, treatment decision, adherence, prognosis). Where suitable reference data relating LDLc concentration to *APOE* genotype are available, such data would not necessarily require direct measurement of traits (e.g., LDLc concentration) but would use national genome-wide clinical databases.

At present, buccal sampling of DNA and genotyping of a specific genetic marker appears to be complex and costly; however, genotyping analysis is available from a single DNA sample for many common polymorphic genetic markers at a cost of $<$0.05p per marker (e.g., http://www.23andMe.com). In the future, population genome-wide data will likely be readily available to calculate GRTIs. As genetic information becomes more widely used in healthcare settings *(10)*, these methods could be widely deployed. The genome-wide genotype data would offer a generic approach, e.g., genotype ratios for some polymorphisms serving as proxies for LDLc concentration and others serving as proxies for HDLc concentration.

The GRTI relies on the availability of genotypes that affect the trait or disease for which a specific intervention takes place. The number of polymorphisms robustly associated with traits or diseases is increasing rapidly (http://www.genome.gov/26525384). These polymorphisms include single-nucleotide polymorphisms associated with lipid subfractions, hypertension, QT interval, other major diseases, behavioral traits, sleep patterns, and other traits that are targets of drug prescribing. The GRTI may be a novel tool for researching the healthcare-delivery process.

LIMITATIONS

Our proof-of-principle was constructed in the context of multiple risk factor management. In this situation, interpreting any differences or similarities between groups in GRTI must consider other risk factors. For example, owing to differences in overall coronary risk, statin prescribing should be more likely in a region of high smoking prevalence, compared with a region of low smoking prevalence. In general, the interpretation of GRTIs depends on the trait, the genotype, and the context of deployment.

Although the major effect of *APOE* genotype is on the LDLc concentration, it also exerts minor effects on HDLc and triglyceride concentrations *(4)*. Across the genotypes arranged in the series ε2/ε2, ε2/ε3, ε2/ε4, ε3/ε3, ε3/ε4, and ε4/ε4, the mean LDLc concentration shows a steep increase, the mean HDLc concentration shows a more modest decrease, and the triglyceride concentration has a U-shaped association. The U shape is mainly attributable to the ε2/ε2 and ε4/ε4 genotypes, but these genotypes were not used in our analyses; hence, any impact of triglycerides on statin prescribing should not affect our findings.

Our sample was restricted to estimating thresholds in women. Greater precision is expected for a larger sample of statin takers. We found no evidence of sample-selection bias by *APOE* genotype; however, "healthy participant" bias might lead to underestimation of differences if the nonparticipants were of a lower socioeconomic class and were underprescribed statins. Although *APOE* genotype frequencies appear to be geographically constant around the UK, a GRTI could not be extrapolated globally because of both clines in allele frequency (Table 1) and differences in diet. Additionally, if there were interactions between genotype and factors other than the trait of interest, the estimates would be biased and would reflect the influence of all of the pleiotropic effects on the trait of interest, rather than a specific threshold. The use of a combination of different genotypes that influence the trait of interest, each of which with a different mode of action and different pleiotropic effects, could mitigate this limitation, however. For example, 95 loci are now known to influence blood lipids *(11)*. In our proof-of-principle study, we used LDLc data estimated from the Friedewald equation, whereas in practice LDLc now would be measured directly; however, our sensitivity analyses based on excluding individuals with triglyceride concentrations >2.5 mmol/L indicated that the principle was robust to indirect estimation of LDLc (see Table 2 in the online Data Supplement). Lastly, any genotype influence on statin initiation or adherence (through a response or side effect) would confound the GRTI. Response may be greater in ε2 carriers *(12)*, but that should not influence adherence.

Author Contributions: *All authors confirmed they have contributed to the intellectual content of this paper and have met the following 3 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; and (c) final approval of the published article.*

Authors' Disclosures or Potential Conflicts of Interest: *Upon manuscript submission, all authors completed the Disclosures of Potential Conflict of Interest form. Potential conflicts of interest:*

Employment or Leadership: G. Davey Smith, MRC Centre for Causal Analyses in Translational Epidemiology; D.A. Lawlor, MRC Centre for Causal Analyses in Translational Epidemiology; I.N. Day, MRC Centre for Causal Analyses in Translational Epidemiology.
Consultant or Advisory Role: None declared.
Stock Ownership: None declared.
Honoraria: None declared.
Research Funding: S. Ebrahim, D.A. Lawlor, Department of Health (England) policy research programme and British Heart Foundation for core support to the British Women's Heart & Health Study. S. Ebrahim, D.A. Lawlor, I.N.M. Day, British Heart Foundation project grant support to develop the BWHHS genetic resource. G. Davey Smith, D.A. Lawlor, I.N.M. Day, UK MRC (G0600705) for MRC Centre for Causal Analyses in Translational Epidemiology. N.M. Davies, MRC Centre for Causal Analyses in Translational Epidemiology PhD studentship. I.N.M. Day, M.R. Abdollahi, University of Bristol start-up funding of Bristol Genetic Epidemiology Laboratory.
Expert Testimony: None declared.

Role of Sponsor: The funding organizations played no role in the design of study, choice of enrolled patients, review and interpretation of data, or preparation or approval of manuscript.

Acknowledgments: Professor Shah Ebrahim, Professor Debbie Lawlor, Professor Peter Whincup, and Dr. Goya Wannamethee co-direct the BWHHS. We thank all staff who collected and entered data for the BWHHS, and we thank all of the general practitioners and their staff who have supported data collection and the women who have participated in the study.

References

1. Davey Smith G, Lawlor DA, Harbord R, Timpson N, Day I, Ebrahim S. Clustered environments and randomized genes: a fundamental distinction between conventional and genetic epidemiology. PLoS Med 2007;4:e352.
2. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007;447:661–78.
3. Evans I, Chalmers I, Thornton H. Testing treatments: better research for better healthcare. London: The British Library; 2006. 224 p.
4. Abdollahi MR, Guthrie PAI, Davey Smith G, Lawlor DA, Ebrahim S, Day INM. Integrated single-label liquid-phase assay of APOE codons 112 and 158 and a lipoprotein study in British women. Clin Chem 2006;52:1420–3.
5. Warnick GR, Knopp RH, Fitzpatrick V, Branson L. Estimating low-density lipoprotein cholesterol by the Friedewald equation is adequate for classifying patients on the basis of nationally recommended cutpoints. Clin Chem 1990;36:15–9.
6. Lawlor DA, Ebrahim S, Davey Smith G. Socioeconomic position in childhood and adulthood and insulin resistance: cross sectional survey using data from British Women's Heart and Health Study. BMJ 2002;325:805.
7. Lawlor DA, Bedford C, Taylor M, Ebrahim S. Geographical variation in cardiovascular disease, risk factors, and their control in older women:

British Women's Heart and Health Study. J Epidemiol Community Health 2003;57:134–40.
8. Liu HC, Hong CJ, Wang SJ, Fuh JL, Wang PN, Shyu HY, Teng EL. ApoE genotype in relation to AD and cholesterol: a study of 2,326 Chinese adults. Neurology 1999;53:962–6.
9. Haddy N, De Bacquer D, Chemaly MM, Maurice M, Ehnholm C, Evans A, et al. The importance of plasma apolipoprotein E concentration in addition to its common polymorphism on inter-individual variation in lipid levels: results from Apo Europe. Eur J Hum Genet 2002;10:841–50.
10. Kaye J. The regulation of direct-to-consumer genetic tests. Hum Mol Genet 2008;17:R180–3.
11. Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, et al. Biological, clinical and population relevance of 95 loci for blood lipids. Nature 2010;466:707–13.
12. Donnelly LA, Palmer CNA, Whitley AL, Lang CC, Doney ASF, Morris AD, Donnan PT. Apolipoprotein E genotypes are associated with lipid-lowering responses to statin treatment in diabetes: a Go-DARTS study. Pharmacogenet Genomics 2008;18:279–87.
13. Humphries SE, Talmud PJ, Hawe E, Bolla M, Day INM, Miller GJ. Apolipoprotein E4 and coronary heart disease in middle-aged men who smoke: a prospective study. Lancet 2001;358:115–9.
14. Ye S, Dunleavey L, Bannister W, Day LB, Tapper W, Collins AR, et al. Independent effects of the −219 G>T and epsilon 2/ epsilon 3/ epsilon 4 polymorphisms in the apolipoprotein E gene on coronary artery disease: the Southampton Atherosclerosis Study. Eur J Hum Genet 2003;11:437–43.
15. Packard CJ, Westendorp RGJ, Stott DJ, Caslake MJ, Murray HM, Shepherd J, et al. Association between apolipoprotein E4 and cognitive decline in elderly adults. J Am Geriatr Soc 2007;55:1777–85.
16. Zhao JH, Brunner EJ, Kumari M, Singh-Manoux A, Hawe E, Talmud PJ, et al. APOE polymorphism, socioeconomic status and cognitive function in mid-life—the Whitehall II longitudinal study. Soc Psychiatry Psychiatr Epidemiol 2005;40:557–63.
17. Cumming AM, Robertson FW. Polymorphism at the apoprotein-E locus in relation to risk of coronary disease. Clin Genet 1984;25:310–3.
18. Taylor AE, Guthrie PAI, Davey Smith G, Golding J, Sattar N, Hingorani AD, et al. IQ, educational attainment, memory and plasma lipids: associations with apolipoprotein E genotype in 5995 children. Biol Psychiatry [Epub ahead of print 2011 Jan 5].