Imputing gene–treatment interactions when the genotype distribution is unknown using case-only and putative placebo analyses—a new method for the Genetics of Hypertension Associated Treatment (GenHAT) study

Barry R. Davis¹,*,†, Charles E. Ford¹, Eric Boerwinkle¹, Donna Arnett², John Eckfeldt³ and Henry Black⁴

¹ University of Texas School of Public Health, Houston, TX 77030, U.S.A.
² University of Minnesota School of Public Health, Minneapolis, MN 55454, U.S.A.
³ Fairview–University Medical Center, Minneapolis, MN 55454, U.S.A.
⁴ Rush-Presbyterian-St. Luke’s Medical Center, Chicago, IL, U.S.A.

SUMMARY

There is a sizeable literature on methods for detecting gene–environment interaction in the framework of case-control studies, particularly with reference to the assumption of independence of genotype and exposure. In the context of a clinical trial, wherein gene–drug interactions with regard to outcomes are examined, these methods may be readily applied, as gene and drug are independent by randomization.

In an active-controlled trial (experimental treatment vs standard) that has collected genotype information, gene–drug interactions can be estimated. In addition, the effect of the experimental treatment vs placebo can be imputed by using data from a historical placebo-controlled trial (standard vs placebo) if either (a) genotype information is available from the historical trial or (b) assumptions are made about the prevalence of genotype and the odds ratios of genotype and disease in the historical trial using information from other studies. Motivation for these procedures is provided by the Genetics of Hypertension Associated Treatment, a large pharmacogenetics, ancillary study of a hypertension clinical trial, and examples from published hypertension trials will be used to illustrate the methods. Copyright © 2004 John Wiley & Sons, Ltd.

KEY WORDS: clinical trial; active control; subgroups; interaction; pharmacogenetics; imputation

INTRODUCTION

Controlled clinical trials are the most rigorous type of study design for evaluating the ability of an intervention to affect an outcome [1]. At any given moment, there are literally hundreds of...
these studies in the United States. Because of the availability of already approved treatments for many common diseases, most trials consider a new or experimental (\(X\)) treatment vs a standard (\(S\)) treatment, in lieu of a placebo (\(P\)).

Opportunistically, large-scale clinical trials provide a useful resource to explore relationships and hypotheses beyond their stated primary and secondary objectives [2]. Many trials have contributed much to our understanding about disease processes through the analysis of possible associations between outcomes and covariates [3–5], including whether certain individuals might benefit more (or less) from a treatment than others [6–8]. One type of covariate that has recently been added to the usual collection of demographic, historical, clinical and laboratory characteristics is genotype [9–11]. A new key question has arisen—'What is the relationship of drugs and genes to outcomes?' or stated another way—'Can we find the right drug for the right person?' [12].

The purpose of this paper is (1) to review the unique features that allow one to leverage the accumulating resources of a large-scale clinical trial by genotyping individuals as the trial progresses and (2) to introduce a new method for estimating gene–treatment (\(X\) vs \(P\)) interaction using information from an active-controlled trial (\(X\) vs \(S\)) and a historical placebo-controlled trial (\(S\) vs \(P\)). Motivation is provided by the Genetics of Hypertension Associated Treatment (GenHAT), a large pharmacogenetics, ancillary study of a hypertension clinical trial. Examples from published hypertension trials will be used to illustrate the method.

The GenHAT study will determine whether variants in hypertension susceptibility genes interact with antihypertensive medication to modify coronary heart disease (CHD) incidence in high-risk hypertensive patients [13]. GenHAT is an ancillary study of the Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT), a double-blind, randomized trial of 42,418 hypertensives, 55 years of age or older, with systolic or diastolic hypertension and one or more risk factors for cardiovascular disease [14,15]. ALLHAT completed follow-up in March 2002. GenHAT is typing variants in multiple hypertension genes, and reporting of results is scheduled for 2004–2005. Analysis of gene–treatment interactions in relation to outcomes will include the primary outcome of ALLHAT, CHD, and the secondary outcomes of stroke, heart failure, blood pressure, and others. To our knowledge, GenHAT is the largest pharmacogenetic study ever conducted. An added strength is its ability to link gene–treatment interactions with important clinical outcomes across diverse ethnic and gender groups.

NOTATION AND MODEL

Let ‘\(D\)’ be an indicator of disease (\(1=\)present, \(0=\)absent), ‘\(E\)’ an indicator of exposure or treatment (\(1=\)test treatment, \(0=\)comparator treatment), and ‘\(g\)’ a genotype at a candidate locus with susceptibility allele ‘\(A\)’ and normal allele ‘\(a\)’. In order to incorporate the effects of candidate genes in a statistical model, it is convenient to assign the possible genotypes into similarly defined indicator variables. For example, we can assume a dominant mode of gene action so that the genotypes \(AA\) and \(Aa\) are equivalent. Then, we can define the genetic covariate \(G(g)=1\) for \(g=aa\) and \(G(g)=0\) for \(g=AA\) or \(g=Aa\). The results we present will generalize to an arbitrary number of genotypes or treatments.

The measure of effect we will use is the odds ratio since this measure for a treatment effect is much more similar in different study designs, populations, and lengths of follow-up than
the use of other measures (e.g. relative risk, risk difference) [27]. As such, the underlying model for relating disease to the genetic and treatment covariates will be logistic regression. The model is given by

\[
P(D = 1 | G, E) = \frac{e^{\alpha + \beta_g G + \beta_t E + \beta_{ge} GE}}{1 + e^{\alpha + \beta_g G + \beta_t E + \beta_{ge} GE}}
\]

and

\[
\text{OR}_g = e^{\beta_g}, \quad \text{OR}_t = e^{\beta_t}, \quad \text{and} \quad \text{OR}_{ge} = e^{\beta_{ge}}
\]

are the genetic, treatment and interaction odds ratios, respectively.

**HEURISTICS**

We will be combining the following methodologies in a novel way—(1) case-only design for testing gene–treatment interactions, (2) putative placebo analysis for combining effect measures (odds ratios) from an experimental vs active-controlled trial and an active vs placebo-controlled trial to compare the experimental treatment to placebo, and (3) imputing gene–treatment interaction in a controlled trial wherein the genotype distributions within treatment groups are unknown. The first two techniques are well known but the third is new. All will be described. The combination of techniques will allow us to impute the gene–treatment interaction for an experimental vs placebo-controlled trial that was never realized.

**CASE-ONLY DESIGN FOR GENOTYPE–TREATMENT INTERACTION**

There is a long tradition in clinical trials of examining treatment-covariate interactions and exploring subgroup effects [18–20]; in epidemiology and genetics there is a growing literature about gene–environment interaction [17, 21–26]. In the present context, environment is treatment and genotype is the covariate. By design, genotype and treatment are independent in clinical trials since the latter is randomly allocated. Theoretically, all baseline characteristics, including genotype will be balanced and therefore the distribution of the genotype will be the same in both (or more) treatment groups. In practice, this is case in almost all large trials (>5000 patients). As a result, there are no special considerations needed for the analysis of genotype–treatment interactions beyond those provided in the references above, except those posed by genotype coding (e.g. dominance) and the assumption of Hardy-Weinberg equilibrium.

Assume that we have genotype information within a two-treatment randomized controlled clinical trial in which genotype has been measured as a dichotomous variable. Due to randomization, there is independence of genotype and treatment (assuming the trial size is large).

Table I shows the expected frequency distribution among participants in the trial according to treatment assignment and genotype [17]. Let SI denote the synergy index, i.e. interaction odds ratio or the ratio of the odds ratio in those with the variant to the odds ratio of those
where \( T = a + b + c + d + e + f + g + h \).

Equations (5) and (6) hold because

\[
\frac{c+d}{\frac{a+b+c+e+g}{a+g+h}} \approx 1 \quad \text{by randomization}
\]
An approximation of the variance of the log of the synergy index for cases can be written as
\[ \frac{1}{a} + \frac{1}{c} + \frac{1}{e} + \frac{1}{g} \] (8)

In traditional case-design gene–environment interaction analyses based on case-control studies, two assumptions are made. One is that gene and environment are independent. The other, a usual one for case-control studies, is that the outcome or disease is rare, say on the order of less than 5 per cent. This is needed to use odds ratios as estimates of risk ratios. In a large randomized clinical trial, gene and treatment are independent by randomization and the assumption of low disease risk is not needed, as the interaction measure (or synergy index) on a multiplicative scale is a ratio of risk ratios (Equation (5)).

**PUTATIVE PLACEBO ANALYSIS**

Active-controlled trials of an experimental treatment vs a standard are usually done because it would be unethical to conduct a placebo-controlled trial if a standard treatment was available. However, in the absence of a placebo, it is difficult to make inferences comparing the new treatment to placebo. Putative placebo methodology is used to make such inferences [27–30]. Key assumptions of this technique are (1) assay sensitivity, i.e. the trial was able to detect differences between treatments and (2) constancy of effect, i.e. the effect of the active control (relative to placebo) was the same as in previous trials [28]. A simple putative placebo analysis is as follows.

Let \( \beta_{SP} \) and \( V_{SP} \) be the log odds ratio and its variance for the standard relative to placebo (from the historical placebo-controlled study) and \( \beta_{XS} \) and \( V_{XS} \) be the log odds ratio and its variance for the experimental treatment relative to the standard (from the current study). Then \( \beta_{XS} + \beta_{SP} \) is an indirect estimate of the log odds ratio estimate of the experimental treatment vs placebo and \( V_{XS} + V_{SP} \) is a pooled estimate variance. A statistic for testing whether the new treatment differs from placebo is \( (\beta_{XS} + \beta_{SP})/\sqrt{V_{XS} + V_{SP}} \) which has a standard normal distribution. The odds ratio and 95 per cent CI are \( e^{\beta_{XS} + \beta_{SP}} \), \( e^{\beta_{XS} + \beta_{SP} \pm 1.96 \sqrt{V_{XS} + V_{SP}}} \), respectively.

**IMPUTING SYNERGY INDEX IN A TRIAL WHERE THE GENOTYPE DISTRIBUTION IS UNKNOWN**

We are given data from a current \( X \) vs \( S \) trial wherein genotype data was collected and a historical \( S \) vs \( P \) trial wherein genotype data was not collected. We know \( \varphi_{XS} \), the odds ratio estimate of disease-genotype relationship in the \( S \) arm of the current \( X \) vs \( S \) where

\[
\varphi_{XS} = \frac{P(D = 1|G = 1; E = S)}{P(D = 0|G = 1; E = S)} \frac{P(D = 1|G = 0; E = S)}{P(D = 0|G = 0; E = S)} \] (9)

In addition, we know, \( s_1 = P(D = 0|E = S) \) and \( s_2 = P(D = 0|E = P) \), the estimates of one minus disease incidence in the two treatment arms.
Suppose for the moment, we had genotype information from the $S$ vs $P$ trial and we consider the following measures

\[ r_1 = P(G = 0|D = 0, E = S) \]  

\[ r_2 = P(G = 0|E = S) \]  

and $\phi_{SP}$, the odds ratio estimate of disease-genotype relationship in the $S$ arm of the $S$ vs $P$ trial where

\[ \phi_{SP} = \frac{P(D = 1|G = 1, E = S)}{P(D = 1|G = 0, E = S)} \frac{P(D = 0|G = 0, E = S)}{P(D = 0|G = 0, E = S)} \]  

Let us consider a second estimate of $\phi_{SP}$ where

\[ \phi_{SP} = \frac{r_1}{(r_2 - r_1 s_1)} \frac{((1 - r_2) - (1 - r_1) s_1)}{(1 - r_1)} \]  

Our algorithm for imputing SI for the $S$ vs $P$ trial is as follows

1. Let $\hat{\phi}_{SP} = \phi_{XS}$ (constancy of genotype effect).
2. Substitute $\hat{r}_2 = P(G = 0)$ for $r_2$, wherein $\hat{r}_2$ is an estimate obtained from a third study (i.e. not the current or historical trial) and is assumed equivalent to $P(G = 0|E = S)$ [$= P(G = 0|E = P)$ by randomization] in the historical $S$ vs $P$ trial.

We can then estimate $r_1$ by $\hat{r}_1 = r_1 \left( \hat{\phi}_{SP}, \hat{r}_2, s_1 \right)$ which is the solution to quadratic equation (13).

\[ \hat{r}_1 = r_1 \left( \hat{\phi}_{SP}, \hat{r}_2, s_1 \right) = 1 + (\hat{r}_2 + s_1)(\hat{\phi}_{SP} - 1) \pm \sqrt{(1 + (\hat{r}_2 + s_1)(\hat{\phi}_{SP} - 1))^2 - 4(s_1(\hat{\phi}_{SP} - 1))(\hat{\phi}_{SP}\hat{r}_2)} \]  

There will always be one solution such that $0 \leq \hat{r}_1 \leq 1$ by the mean value theorem. Also, we assume that $P(G = 0|D = 0, E = S) = P(G = 0|D = 0, E = P)$ (or $\hat{r}_1$ is the same in the standard and placebo arms). This is true if $P(D = 1|E = S)$ and $P(D = 1|E = P)$ are small (rare disease assumption) and thus

\[ P(G = 0|E = S) = P(G = 0|D = 0, E = S)P(D = 0|E = S) \]  

\[ + P(G = 0|D = 1, E = S)P(D = 1|E = S) \]  

or

\[ P(G = 0|E = S) \approx P(G = 0|D = 0, E = S) \]  

Similar reasoning applies for $P(G = 0|D = 0, E = P)$, and since $P(G = 0|E = S) = P(G = 0|E = P)$ by randomization, then $P(G = 0|D = 0, E = S) = P(G = 0|D = 0, E = P)$. Note that
although before we did not need to make the rare disease assumption for estimating gene–
treatment interaction using cases only in the context of a clinical trial, we have to now use
this assumption to carry out our imputation.

Completing Table III, the case-only SI for the $S$ vs $P$ trial is

$$SI_{SP} = \frac{((1 - \hat{r}_2) - (1 - \hat{r}_1)s_1)s_{SP}}{(\hat{r}_2 - \hat{r}_1s_1)s_{SP}} \times \frac{(\hat{r}_2 - \hat{r}_1s_2)P_{SP}}{((1 - \hat{r}_2) - (1 - \hat{r}_1)s_2)P_{SP}}$$

$$= \frac{((1 - \hat{r}_2) - (1 - \hat{r}_1)s_1)}{(\hat{r}_2 - \hat{r}_1s_1)} \times \frac{(\hat{r}_2 - \hat{r}_1s_2)}{((1 - \hat{r}_2) - (1 - \hat{r}_1)s_2)}$$

(16)

where $\hat{\phi}_{SP}$, $\hat{r}_2$, $s_1$, and $s_2$ are all independently estimated and $S_{SP}$ and $P_{SP}$ are the number of participants in the standard and placebo arms of the placebo-controlled trial.

The naïve variance of $\ln SI_{SP}$ is

$$\frac{1}{(\hat{r}_2 - \hat{r}_1s_1)s_{SP}} + \frac{1}{((1 - \hat{r}_2) - (1 - \hat{r}_1)s_1)s_{SP}} + \frac{1}{(\hat{r}_2 - \hat{r}_1s_2)P_{SP}}$$

$$+ \frac{1}{((1 - \hat{r}_2) - (1 - \hat{r}_1)s_2)P_{SP}}$$

(17)

The actual variance can be obtained using the delta method. It is

$$\left(\frac{d}{d\hat{\phi}_{SP}} [\ln (SI_{SP})]\right)^2 \text{Var}(\hat{\phi}_{SP}) + \left(\frac{d}{d\hat{r}_2} [\ln (SI_{SP})]\right)^2 \text{Var}(\hat{r}_2) + \left(\frac{d}{ds_1} [\ln (SI_{SP})]\right)^2 \text{Var}(s_1)$$

$$+ \left(\frac{d}{ds_2} [\ln (SI_{SP})]\right)^2 \text{Var}(s_2)$$

(18)

where $\text{Var}(\hat{\phi}_{SP})$, $\text{Var}(\hat{r}_2)$, $\text{Var}(s_1)$, and $\text{Var}(s_2)$ are calculable using data from the three
sources—the $X$ vs $S$ trial, the study used to obtain the genotype distribution, and the $S$
vs $P$ trial. The actual variance can be smaller than the naïve one if the former two studies’
sample sizes are larger than the one obtained by summing the denominators of Equation (17).

The implications of this are that one needs to evaluate both the clinical significance of the
point estimate and its confidence interval in the context of the sample sizes of all studies
employed for this procedure.

**IMPUTING GENE–TREATMENT INTERACTION IN AN UNREALIZED $X$ VS $P$
TRIAL—ASSUMPTIONS**

We will present two methods imputing gene–treatment interaction in an unrealized $X$ vs $P$
trial. The key assumptions of both methods are assay sensitivity, and constancy of treatment
and genotype effects. Assay sensitivity means that the trial was able to detect differences be-
tween treatments. This is self-documented in superiority trials. However, the following factors
can reduce assay sensitivity—poor compliance with therapy, poor responsiveness of the study population, concomitant medications, a population that improves spontaneously, poor diagnostic criteria, inappropriate measures of drug effect, excessive variability of measurements and biased assessment of endpoints. All of these factors for most trials can be assessed. Constancy of treatment effect means that the effect of the active standard ($S$) [relative to placebo ($P$)] in previous trial applies today. Similarly, constancy of genotype effect means that the effect of genotype applies today. Factors that reduce con/fidence in these assumptions are inclusion/exclusion criteria, endpoint definitions, dosing regimen, and concomitant medications. These can be assessed and qualitatively (but not quantitatively) aid in the overall interpretation.

**Method 1—the putative placebo method**

We have the following data available to us—(1) cases by treatment and genotype from the $X$ vs $S$ trial, and (2) cases by treatment and genotype from the $S$ vs $P$ trial. The algorithm is to calculate the log of the case-only SI and its variance using data in (1) and (2) and sum the respective estimates. The final ‘putative placebo’ estimate of the case-only SI can and its 95 per cent CI are (following the notation of Table II)

\[
\frac{a_{XS}g_{XS}}{c_{XS}e_{XS}} \times \text{exp}\left(\pm 1.96 \sqrt{\frac{1}{a_{XS}} + \frac{1}{c_{XS}} + \frac{1}{e_{XS}} + \frac{1}{g_{XS}} + \frac{1}{a_{SP}} + \frac{1}{c_{SP}} + \frac{1}{e_{SP}} + \frac{1}{g_{SP}}}\right)
\]

**Method 2—The imputed synergy index plus putative placebo method**

If we do not have genotype information available from the $S$ vs $P$ trial, we can impute an SI based on the method outlined under the section ‘imputing synergy index in a trial where the genotype distribution is unknown’. Using the imputed SI plus the calculated one from the $X$ vs $S$ trial, the putative case-only SI for the $X$ vs $P$ comparison would be

\[
\text{SI}_{XP} = \frac{a_{XS}g_{XS}}{c_{XS}e_{XS}} \frac{((1 - \hat{r}_2) - (1 - \hat{r}_1)s_1)}{(\hat{r}_2 - \hat{r}_1)s_1} \frac{(\hat{r}_2 - \hat{r}_1)s_2}{((1 - \hat{r}_2) - (1 - \hat{r}_1)s_2)}
\]

\[(19)\]
Table III. 2 × 2 × 2 of historical placebo-controlled trial with imputed values for cells.

<table>
<thead>
<tr>
<th></th>
<th>Standard treatment</th>
<th></th>
<th>Placebo treatment</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>G = 0</td>
<td>G = 1</td>
<td>Total</td>
<td>G = 0</td>
</tr>
<tr>
<td>D = 0</td>
<td>( \tilde{r}<em>1 s_1 S</em>{Sp} )</td>
<td>(1 - ( \tilde{r}<em>1 )) s_1 S</em>{Sp}</td>
<td>s_1 S_{Sp}</td>
<td>D = 0</td>
</tr>
<tr>
<td>D = 1</td>
<td>(( \tilde{r}_2 - \tilde{r}<em>1 )) s_1 S</em>{Sp}</td>
<td>((1 - ( \tilde{r}_2 )) - (1 - ( \tilde{r}<em>1 )) s_1) S</em>{Sp}</td>
<td>(1 - s_1) S_{Sp}</td>
<td>D = 1</td>
</tr>
<tr>
<td>Total</td>
<td>( \tilde{r}<em>2 S</em>{Sp} )</td>
<td>(1 - ( \tilde{r}<em>2 )) S</em>{Sp}</td>
<td>S_{Sp}</td>
<td>( \tilde{r}<em>2 P</em>{Sp} )</td>
</tr>
</tbody>
</table>

Where \( S_{Sp} \) and \( P_{Sp} \) are the number of participants in the standard and placebo arms of the placebo-controlled trial.
The naïve variance of $\ln SI_{XP}$ is

$$
\frac{1}{(\hat{r}_2 - \hat{r}_1s_1)SP} + \frac{1}{((1 - \hat{r}_2) - (1 - \hat{r}_1)s_1)SP} + \frac{1}{(\hat{r}_2 - \hat{r}_1s_2)SP} + \frac{1}{((1 - \hat{r}_2) - (1 - \hat{r}_1)s_2)SP}
+ \frac{1}{aXS} + \frac{1}{eXS} + \frac{1}{cXS} + \frac{1}{gXS}
$$

(20)

The actual variance can be smaller for the same reason previously described. It is

$$
\left(\frac{d}{d\phi_{XS}} [\ln(SI_{AP})]\right)^2 \text{Var}(\phi_{XS}) + \left(\frac{d}{d\hat{r}_2} [\ln(SI_{AP})]\right)^2 \text{Var}(\hat{r}_2) + \left(\frac{d}{ds_1} [\ln(SI_{AP})]\right)^2 \text{Var}(s_1)
+ \left(\frac{d}{ds_2} [\ln(SI_{AP})]\right)^2 \text{Var}(s_2) + \frac{1}{aXS} + \frac{1}{eXS} + \frac{1}{cXS} + \frac{1}{gXS}
$$

(21)

EXAMPLES

Details of all examples using Method 2 are presented in Table IV. As we do not have genotype information yet from GenHAT, we shall use a well-known genetic marker, sex (men vs women) obtained from the ALLHAT study. The interaction odds ratio for hospitalized/fatal heart failure in ALLHAT for doxazosin vs chlorthalidone and men vs women was 1.19, 95 per cent CI (0.91, 1.56) or the odds of developing heart failure on doxazosin compared with chlorthalidone was 19 per cent higher for men than for women.

There was no placebo in ALLHAT. Using data from the Systolic Hypertension in the Elderly Program (SHEP) [31,32], a placebo-controlled (chlorthalidone vs placebo) hypertension trial, the interaction odds ratio [chlorthalidone vs placebo and men vs women] was 1.21, 95 per cent CI (0.58,2.52), i.e. or the odds of developing heart failure on chlorthalidone compared with placebo was 21 per cent higher for men than for women.

The putative placebo method then yields a synergy index [doxazosin vs placebo and men vs women] of 1.44, 95 per cent CI (0.65,3.15), or the odds of developing heart failure on doxazosin compared with placebo was 44 per cent higher for men than for women.

Using the second method, we calculated the odds ratio of the heart failure outcome and sex in the chlorthalidone group within ALLHAT ($\phi_{SP} = 1.16$), estimated the proportion of women in SHEP ($\hat{r}_2 = 0.57$) [31], and used the estimates of the complement of the outcome incidences in the active standard and placebo groups in SHEP ($s_1 = 0.981, s_2 = 0.967$) [32]. Then the imputed SI in SHEP [chlorthalidone vs placebo and men vs women] for the outcome was 1.07, 95 per cent CI (0.98,1.16), or the odds of developing heart failure on chlorthalidone compared with placebo was 7 per cent higher for men than for women. The synergy index [doxazosin vs placebo and men vs women] was 1.27, 95 per cent CI (0.96–1.68), or the odds of developing heart failure on doxazosin compared with placebo is 27 per cent higher for men than for women.

A recently published article by Psaty and others [33] examined the relationship of alpha-adducin gene with myocardial infarction (MI) and stroke. The synergy index [other antihy-
Table IV. Synergy indices (experimental vs standard, standard vs placebo, experimental vs placebo) using method 2.

<table>
<thead>
<tr>
<th>Study and outcome</th>
<th>Treatment (experimental vs standard) — genotype interaction</th>
<th>Imputed SI Naive 95 per cent CI</th>
<th>Treatment (standard vs placebo) — genotype interaction</th>
<th>Method 2 SI Naive 95 per cent CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALLHAT Hosp/fatal heart failure</td>
<td>Doxazosin/chlorthalidone — men/women</td>
<td>1.19 (0.91, 1.56)</td>
<td>Chlorthalidone/Placebo — men/women</td>
<td>1.27 (0.58, 2.78)</td>
</tr>
<tr>
<td>Psaty (33) Myocardial infarction</td>
<td>Others/diuretic-adducin genotype (variant/wild type)</td>
<td>2.13 (1.16, 3.85)</td>
<td>Chlorthalidone/Placebo-Adducin Genotype (Variant/Wild type)</td>
<td>2.47 (1.01, 6.09)</td>
</tr>
<tr>
<td>Psaty (33) Stroke</td>
<td>Others/diuretic-adducin genotype (variant/wild type)</td>
<td>2.47 (0.76, 3.45)</td>
<td>Chlorthalidone/Placebo-Adducin Genotype (Variant/Wild type)</td>
<td>1.70 (0.67, 4.34)</td>
</tr>
</tbody>
</table>

* Actual treatment (standard vs placebo)-subgroup interaction SI and 95 per cent CI is 1.21 (0.52, 2.52).
† Method 1 estimate of treatment (new vs placebo) subgroup-interaction SI and 95 per cent CI is 1.44 (0.65, 3.15).
pertensive therapy vs diuretic and adducin variant vs adducin wild type] for MI was 2.13, 95 per cent CI (1.16, 3.85), or the odds of having an MI on other antihypertensive therapy compared with diuretic was 113 per cent higher for the variant than the wild type. Using the second method, we calculated the odds ratio of the MI outcome and adducin variant in the diuretic group from Psaty et al. [33] (\(\hat{\phi}_{SP} = 1.61\)), estimated the proportion of those without the adducin variant in SHEP (\(\hat{r}_2 = 0.75\)) [34], and used the estimates of the complement of the outcome incidences in the active standard and placebo groups in SHEP (\(s_1 = 0.973, s_2 = 0.959\)). Then, the imputed synergy index in SHEP [diuretic (chlorthalidone) and adducin variant vs adducin wild type] was 1.17, 95 per cent CI (0.97, 1.42), or the odds of having an MI on diuretic compared with placebo was 17 per cent higher for the variant than the wild type. We note here that we used chlorthalidone as representative of the class of diuretics. Large controlled trials have shown similar mortality or morbidity reductions with many different types of diuretics [35].

The second method for the putative SI [other antihypertensive therapy vs placebo] yielded an SI of 2.47, 95 per cent CI (1.32, 4.64), or the odds of having an MI on other antihypertensive therapy compared with placebo was 147 per cent higher for the variant than the wild type. This implies that, for those with the adducin variant compared to those with the adducin wild type, being on other antihypertensive therapies (instead of diuretic) placed one at increased risk for an MI.

In a third example, the odds of having a stroke on other antihypertensive therapy compared with diuretic was 61 per cent higher for the adducin variant than the wild type: SI = 1.61, 95 per cent CI (0.76, 3.45). Using the second method, we calculated the odds ratio of the stroke outcome and adducin variant in the diuretic group from Psaty et al. [33] (\(\hat{\phi}_{SP} = 1.16\)), estimated the proportion of those with the adducin variant in SHEP (\(\hat{r}_2 = 0.25\)) [34], and used the estimates of the complement of the outcome incidences in the active standard and placebo groups in SHEP (\(s_1 = 0.956, s_2 = 0.933\)) [31]. Then, the imputed synergy index in SHEP [diuretic (chlorthalidone) and adducin variant vs adducin wild type] was 1.05, 95 per cent CI (0.86, 1.29) or the odds of having a stroke on other diuretic compared with placebo was 5 per cent higher for the variant type than the wild type. The second method for the putative SI [other antihypertensive therapy vs placebo] yielded an SI of 1.70, 95 per cent CI (0.78, 3.71) or the odds of having a stroke on other antihypertensive therapy compared with placebo was 70 per cent higher for the variant than the wild type. This implies that, for those with the adducin variant compared to those with the adducin wild type, being on other antihypertensive therapies conferred a similar stroke risk to being on a diuretic.

**DISCUSSION**

Simultaneous genotyping within the framework of ongoing large clinical trials facilitates our understanding of the biologic underpinning of interindividual response to treatments. Questions such as ‘Are there subgroups of individuals, at least partially defined by their genotype, who would benefit more from one type of treatment than another?’ can be addressed. Specifically, we can examine gene–treatment interactions with regard to clinical outcomes and, in addition, indirectly examine such interactions in similar trials that used the same agent (or same class of agent) without necessarily having genetic information from these trials. Presently, there is
an increasing shift in emphasis from single genes to multiple genes operating in pathways. Further work on examining gene–gene and gene–gene–treatment interactions in clinical trials is required.

In the examples presented, we can examine the three assumptions—assay sensitivity, constancy of treatment effect and constancy of genotype effect—used to calculate gene–treatment interactions in an unrealized trial. For the heart failure example, treatment differences were detected in ALLHAT and SHEP, but there were no significant sex–treatment interactions in either trial. Both trials were double-blind, randomized and well executed. It is likely that there were no sex–treatment interactions and these were not detectable. Also, both studies had similar endpoint definitions [36, 37], and dosing regimens for diuretic (chlorthalidone) and step-up drugs (atenolol, reserpine). However, there were differences in inclusion/exclusion criteria so that not all SHEP participants would have qualified for ALLHAT and not all ALLHAT participants would have qualified for SHEP. In addition, many of the concomitant medications used in ALLHAT were either infrequently used or not used at all in SHEP, as the former trial ran from 1994 through 2002 while the latter ran from 1985 through 1991.

For the MI and stroke examples, treatment differences were not detected in the case-control study but were detected in SHEP. A gene–treatment interaction was detected in the case-control study for MI but not for stroke, and no gene–treatment interactions were detected in SHEP using method 2. Both studies had similar endpoint definitions, but the dosing regimens were not the same, as the case-control study included all diuretics (thiazide and loop) whereas SHEP essentially used chlorthalidone (a thiazide diuretic). The participants in the two studies were somewhat different (e.g. more African–Americans and fewer persons with diabetes in SHEP). The case-control study used cases from the period 1995 through 1998 compared with SHEP’s earlier time frame, so it is likely the concomitant drugs would have been different.

Directions for future research on the present methods include (1) using meta-analyses rather than single trials for imputation and (2) imputing $P(G)$ in the placebo-controlled trial from a logistic model in active-controlled trial wherein the dependent variable is genotype and independent variables are characteristics in common between the two studies. In addition, more formal work is needed on verifying the assumptions that are used for both the placebo putative analyses and the present methodology.

Given the continuing identification of new genes and their relation to disease processes, the techniques set forth in this paper allow one to impute a gene–treatment interaction in trials wherein no genotype information was collected. This would provide an enormous resource for designing new studies with specifically targeted populations. In addition, although we specifically examined gene–treatment interactions, the method for imputing a synergy index can be used for other types of subgroups. For example, in the ALLHAT study wherein doxazosin (an alpha-blocker) was compared to chlorthalidone (a diuretic), a significantly higher rate of heart failure was observed in those participants assigned to doxazosin. One possible explanation was that before entering the trial, many individuals had been on a diuretic. Perhaps, the discontinuation of all previous antihypertensive medications placed individuals who had subclinical heart failure at risk. Those who went back on a diuretic were not at risk while those who went on doxazosin were. This hypothesis could have been examined if type of prior medications had been collected in ALLHAT. Unfortunately, it was not. However, if such information was available in another hypertension trial that used either a diuretic or alpha-blocker, the technique explored in this article could be applied to answer the question.
In conclusion, pharmacogenetics presents many exciting opportunities to understand disease mechanisms and improve the health of individuals with targeted treatments. We have proposed a new tool that should add to the growing possibilities of future research. In our own case, we will apply the method to GENHAT, whose results will be forthcoming in the next 2 years. The derived information may prove to be very important in guiding treatment decisions.

REFERENCES