# Trend tests for genetic association using population-based cross-sectional complex survey data

DEWEI SHE

*Department of Statistics, The George Washington University, Washington, DC 20052, USA and
The EMMES Corporation, 401 North Washington Street, Rockville, MD 20850, USA*

YAN LI

*Department of Mathematics, University of Texas at Arlington, Arlington, TX 76019, USA*

HONG ZHANG

*Department of Statistics and Finance, University of Science and Technology of China,
Hefei, Anhui 230026, People's Republic of China and Biostatistics Branch, Division of Cancer
Epidemiology and Genetics, National Cancer Institute, Rockville, MD 20852, USA*

BARRY I. GRAUBARD

*Biostatistics Branch, Division of Cancer Epidemiology and Genetics,
National Cancer Institute, Bethesda, MD 20892, USA*

ZHAOHAI LI\*

*Department of Statistics, The George Washington University, Washington, DC 20052, USA and
Biostatistics Branch, Division of Cancer Epidemiology and Genetics,
National Cancer Institute, Rockville, MD 20852, USA*
zli@gwu.edu

SUMMARY

Genetic data collected from surveys such as the Third National Health and Nutrition Examination Survey (NHANES III) enable researchers to investigate the association between wide varieties of health factors and genetic variation for the US population. Tests for trend in disease with increasing number of alleles have been developed for simple random samples. However, surveys such as the NHANES III have complex sample designs involving multistage cluster sampling and sample weighting. These types of sample designs can affect Type I error and power properties of statistical tests based on simple random samples. In order to address these issues, we have derived tests of trend based on Wald and quasi-score statistics, with and without assuming a genetic model, that account for the complex sampling design. The finite-sample properties of the proposed test procedures are evaluated via Monte Carlo simulation studies. We make recommendations about the choice of the test statistic depending on whether or not the underlying genetic model is known. Proposed test statistics are applied to NHANES III data to test for associations

---

\*To whom correspondence should be addressed.

between the locus ADRB2 (rs1042713) and obesity, between VDR (rs2239185) and high blood lead level, and between TGFB1 (rs1982073) and asthma.

## 1. INTRODUCTION

The Third National Health and Nutrition Examination Survey (NHANES III) provides national cross-sectional estimates of the health and nutritional status of the US civilian noninstitutionalized population. During the second phase of NHANES III, blood lymphocytes were collected from 7159 participants aged 12 years and older in anticipation of advances in genetic research. Linkage of the NHANES III phenotype data with this genetic information provides the opportunity to investigate the association of a wide variety of health factors with regard to genetic variations at the US population level (National Center for Health Statistics, 2008a).

National surveys like the NHANES III employ complex sampling plans that have stratified multistage cluster sample designs, which add at least 2 complexities to the data analyses. First, cluster sampling may induce correlation between individual observations within sampled clusters. The standard errors can be underestimated if this correlation is ignored. Second, varying selection probabilities used to sample individuals produce different sample weights for surveyed subjects. If these sample weights are correlated with the characteristics of research interest for the observations, then an analysis that does not take this into account can be biased (Korn and Graubard, 1999). Test procedures developed for simple random samples are generally unsuitable for the analysis of data from these complex sample designs.

For simple random samples, the Cochran–Armitage trend test (Armitage, 1955) or efficient score trend test has been used to evaluate the association between a candidate gene and a disease using a case–control design (Morton and Collins, 1998; Slager and Schaid, 2001; Freidlin *and others*, 2002; Zheng and Gastwirth, 2006; Epstein *and others*, 2007). Recently, Ryckman *and others* (2008) proposed a prevalence-based association test for a case–control study. These tests can also be applied to cross-sectional studies with simple random samples such as surveys where prevalent disease cases and nondisease (controls) individuals are sampled together. When these tests are applied to the data from cross-sectional surveys that have complex sample designs, they need to be modified to take account of the sample design. Our objective in this paper was to develop appropriate test procedures for the association between a candidate gene and disease in surveys that have complex samples.

## 2. METHODS

### 2.1 *Cochran–Armitage trend test*

Under simple random sampling, the association of a candidate gene with a disease can be evaluated with the Cochran–Armitage trend test. Assuming that allele $A$ of the candidate gene is the allele of high risk and allele $a$ is any of the other alleles, the genotype data obtained from a case–control study or a cross-sectional study can be represented as the following: $r_i$ and $s_i$ for $i = 0, 1$, and 2 are the sizes of cases and controls with the number $i$ of allele $A$, $r$ and $s$ are the sample sizes of cases and controls, and $r + s = n$. Given the case and control status, $(r_0, r_1, r_2)^T$ and $(s_0, s_1, s_2)^T$ follow trinomial distributions with parameter vectors of $(p_0, p_1, p_2)^T$ and $(q_0, q_1, q_2)^T$ for cases and controls, respectively. Denote $K$ as the disease prevalence, $\mathbf{f} = (f_0, f_1, f_2)^T$ as the penetrances of genotypes ($aa, aA, AA$), and $\gamma = \left(1, \frac{f_1}{f_0}, \frac{f_2}{f_0}\right)^T = (1, \gamma_1, \gamma_2)^T$ as

the relative risks. We have $K = \sum_{i=0}^{2} f_i g_i = f_0 \sum_{i=0}^{2} \gamma_i g_i$, $p_i = \frac{f_i g_i}{K}$, and $q_i = \frac{(1-f_i)g_i}{1-K}$ for $i = 0, 1$, and 2, where $\mathbf{g} = (g_0, g_1, g_2)^T$ are the population genotype frequencies (Slager and Schaid, 2001; Freidlin *and others*, 2002; Zheng and Gastwirth, 2006).

Let $\mathbf{x} = (x_0, x_1, x_2)^T$ be the score vector assigned to genotypes $(aa, aA, AA)$ with $\mathbf{x} = (0, 0.5, 1)^T$ for additive model, $\mathbf{x} = (0, 0, 1)^T$ for recessive model, and $\mathbf{x} = (0, 1, 1)^T$ for dominant model (Sasieni, 1997; Freidlin *and others*, 2002), then the Cochran–Armitage trend test statistic for the null hypothesis of no association between the candidate gene and the disease of interest, that is, $H_0$: $f_0 = f_1 = f_2$, can be written as

$$T_T = \frac{U_T}{\sqrt{\widehat{\text{Var}}(U_T)}} \tag{2.1}$$

or $T_T^2$, where $U_T = \frac{1}{n} \sum_{i=0}^{2} x_i(sr_i - rs_i)$ and $\widehat{\text{Var}}(U_T) = \frac{s^2}{n^2} \widehat{\text{Var}}\left(\sum_{i=0}^{2} x_i r_i\right) + \frac{r^2}{n^2} \widehat{\text{Var}}\left(\sum_{i=0}^{2} x_i s_i\right)$. $T_T$ has an asymptotic normal distribution with mean 0 and variance 1, and $T_T^2$ has an asymptotic $\chi_1^2$ distribution under the null hypothesis (Sasieni, 1997; Slager and Schaid, 2001; Freidlin *and others*, 2002; Zheng and Gastwirth, 2006).

Surveys such as NHANES III are household surveys that have a stratified multistage sample designs, where primary sample units (PSUs), for example, counties, are sampled from strata at the first stage of sampling and additional stages of sampling are conducted within the sampled PSUs. At the last stage, individuals are sampled from sampled households (for details about the NHANES III sample design, see Ezzati *and others*, 1992). For each individual, the inverse of the product of the selection probabilities across all the stages of sampling is their sample weight. Define the weighted analogies of $n$, $r$, $s$, $r_i$, $s_i$, and $n_i$ as $n_w = \sum_{j=1}^{n} w_j$, $r_w = \sum_{j=1}^{n} w_j y_j$, $s_w = \sum_{j=1}^{n} w_j(1 - y_j)$, $r_{iw} = \sum_{j=1}^{n} w_j y_j G_j^i$, $s_{iw} = \sum_{j=1}^{n} w_j(1 - y_j)G_j^i$, and $n_{iw} = r_{iw} + s_{iw}$, where $w_j$ is the sample weight associated with the $j$th individual, $y_j = 1$ if the $j$th individual is case and 0 otherwise, and $G_j^i = 1$ if the $j$th individual has the number $i$ of allele $A$ at the locus of interest. The Cochran–Armitage trend test statistic for the null hypothesis can be formed as

$$T_{Tw} = \frac{U_{Tw}}{\sqrt{\widehat{\text{Var}}(U_{Tw})}} \tag{2.2}$$

or $T_{Tw}^2$, where $U_{Tw} = \frac{1}{n_w} \sum_{i=0}^{2} x_i(s_w r_{iw} - r_w s_{iw}) = \sum_{i=0}^{2} x_i\left(r_{iw} - \frac{r_w}{n_w} n_{iw}\right)$.

Denoting $z_j = \frac{\partial U_{Tw}}{\partial w_j}$, $U_{Tw}$ can be approximated by $\sum_{j=1}^{n} w_j z_j$ (Shah, 2004). An estimate of the variance of $U_{Tw}$ is

$$\widehat{\text{Var}}(U_{Tw}) = \sum_{h=1}^{H} \frac{m_h}{m_h - 1} \sum_{l=1}^{m_h} (z^{(hl)} - \bar{z}^{(h)})^2, \tag{2.3}$$

where $z^{(hl)}$ is the weighted sum of $z_j$ in the $l$th PSU of the $h$th strata, $\bar{z}^{(h)} = \sum_{l=1}^{m_h} \frac{z^{(hl)}}{m_h}$, $H$ is the total number of strata, and $m_h$ is the number of PSUs within $h$th stratum (Korn and Graubard, 1999). Under the null hypothesis, $T_{Tw}$ has an asymptotic normal distribution with mean 0 and variance 1, and $T_{Tw}^2$ has an asymptotic $\chi_1^2$ distribution (Graubard and Korn, 1993). When $f = \sum_{h=1}^{H} m_h - H$ is not large, an F-version of the test statistic $FT_{Tw}^2 = \frac{f-1+1}{f} T_{Tw}^2$ has an asymptotic $F$ distribution with degrees of freedom 1 and $f$ under $H_0$ (Korn and Graubard, 1999; Li and Graubard, 2009).

$U_{Tw}$ can be expressed as $\sum_{j=1}^{n} w_j x_j\left(y_j - \frac{r_w}{n_w}\right)$, which is the pseudo-score function based on logistic regression model. The test procedure proposed above is the same as the quasi-score test for a simple linear logistic regression model proposed by Rao *and others* (1998).

## 2.2 *Quasi-efficient score test*

Under simple random sampling, the log-likelihood function can be written as

$$\ell = \sum_{i=0}^{2} r_i \log(p_i) + \sum_{i=0}^{2} s_i \log(q_i) + \text{constant}$$

$$= \sum_{i=0}^{2} r_i \log(f_i) + \sum_{i=0}^{2} s_i \log(1 - f_i) - r \log \left( \sum_{i=0}^{2} f_i g_i \right) - s \log \left\{ \sum_{i=0}^{2} (1 - f_i) g_i \right\} + \text{constant}.$$

In the following, we develop quasi-efficient score test with mode of inheritance known or unknown.

*Mode of inheritance is known.* With simple random sampling, the null hypothesis $H_0$, $f_0 = f_1 = f_2$, can be tested using a score test statistic

$$T_{S1} = \frac{U_{S1}}{\sqrt{\widehat{\text{Var}}(U_{S1})}} \tag{2.4}$$

or $T_{S1}^2$, where $U_{S1} = \frac{1}{f_0}\{(r_2 - rg_2) + x_1(r_1 - rg_1)\} - \frac{1}{1-f_0}\{(s_2 - sg_2) + x_1(s_1 - sg_1)\}$ is a weighted score, $\widehat{\text{Var}}(U_{S1}) = \frac{1}{f_0^2}\{r_2 + x_1^2 r_1 - r(g_2 + x_1 g_1)^2\} + \frac{1}{(1-f_0)^2}\{s_2 + x_1^2 s_1 - s(g_2 + x_1 g_1)^2\}$ is an estimate of the variance, and score $x_1$ equals 0 for recessive inheritance model, $\frac{1}{2}$ for additive model, and 1 for dominant model. The derivation is similar to that given by Li *and others* (2005). $T_{S1}$ has an asymptotic normal distribution with mean 0 and variance 1, and $T_{S1}^2$ has an asymptotic $\chi_1^2$ distribution under the null hypothesis.

In a complex sampling setting, the quasi-efficient score test statistic for the null hypothesis can be formed as

$$T_{S1w} = \frac{U_{S1w}}{\sqrt{\widehat{\text{Var}}(U_{S1w})}}, \tag{2.5}$$

where $U_{S1w}$ is the weighted analogy of $U_{S1}$, which can be derived from a pseudo weighted likelihood approach (Rao *and others*, 1998). To estimate $\text{Var}(U_{S1w})$, we first approximate $U_{S1w}$ by its Taylor expansion (Shah, 2004). Standard methods for survey sampling can then be used to estimate the variance, which has an expression similar to (2.3). When $f = \sum_{h=1}^{H} m_h - H$ is not large, an F-version of the test statistic $FT_{S1w}^2 = \frac{f-1+1}{f} T_{S1w}^2$ has an asymptotic $F$ distribution with degrees of freedom 1 and $f$ under $H_0$ (Korn and Graubard, 1999; Li and Graubard, 2009).

*Mode of inheritance is unknown.* For simple random sample, the score test statistic of the null hypothesis can be expressed as

$$T_{S2}^2 = U^T I^{-1} U, \tag{2.6}$$

where $U = \left( \frac{\partial \ell}{\partial f_2}|_{f_2=f_1=f_0}, \frac{\partial \ell}{\partial f_1}|_{f_2=f_1=f_0} \right) = \left( \frac{1}{f_0}(r_2 - rg_2) - \frac{1}{1-f_0}(s_2 - sg_2), \frac{1}{f_0}(r_1 - rg_1) - \frac{1}{1-f_0}(s_1 - sg_1) \right)$ and

$$I = \begin{pmatrix} \frac{\partial^2 \ell}{\partial f_2^2}|_{f_2=f_1=f_0} & \frac{\partial^2 \ell}{\partial f_2 \partial f_1}|_{f_2=f_1=f_0} \\ \frac{\partial^2 \ell}{\partial f_1 \partial f_2}|_{f_2=f_1=f_0} & \frac{\partial^2 \ell}{\partial f_1^2}|_{f_2=f_1=f_0} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1}{f_0^2}(r_2 - rg_2^2) + \frac{1}{(1-f_0)^2}(s_2 - sg_2^2) & -\frac{rg_1g_2}{f_0^2} - \frac{sg_1g_2}{(1-f_0)^2} \\ -\frac{rg_1g_2}{f_0^2} - \frac{sg_1g_2}{(1-f_0)^2} & \frac{1}{f_0^2}(r_1 - rg_1^2) + \frac{1}{(1-f_0)^2}(s_1 - sg_1^2) \end{pmatrix}.$$

$T_{S2}^2$ has an asymptotic $\chi_2^2$ distribution under the null hypothesis.

In a complex sampling setting, a quasi-score test statistic $T_{S2w}^2$ is obtained by incorporating the sample weights in $T_{S2}^2$. This can be done because $T_{S2}^2$ is a quadratic test statistic (Graubard and Korn, 1993), that is, $T_{S2w}^2 = U_w^T I_w^{-1} U_w$, where $U_w$ and $I_w$ are the weighted analogies of $U$ and $I$. $T_{S2w}^2$ does not have an asymptotic chi-square distribution because of correlation induced by cluster sampling and the sampling weights. Graubard and Korn (1993) showed that (i) $\sqrt{n} U_w \xrightarrow{D} N(0, \Sigma)$ and (ii) $T_{S2w}^2 \xrightarrow{D} \bar{\lambda} \chi^2(2)$, where $\bar{\lambda}$ is the average eigenvalue of $\Sigma \Delta$ and $\frac{1}{n} I_w^{-1} \xrightarrow{P} \Delta$.

Variance–covariance of $U_w$ can be estimated using standard methods from survey sampling (Korn and Graubard, 1999). Denoting $\hat{\Sigma}$ as the estimated variance, we propose a Wald test statistic as the following: $T_{S2wald}^2 = U_w^T \hat{\Sigma}^{-1} U_w$, which is asymptotically distributed as $\chi_2^2$ under $H_0$. When $f = \sum_{h=1}^H m_h - H$ is not large, an $F$-version of the Wald statistic, $FT_{S2wald}^2 = (f - 2 + 1)/(2f) T_{S2wald}^2$, which has an asymptotic $F_{2, f-2+1}$ distribution under $H_0$, is often used instead (Korn and Graubard, 1999; Li and Graubard, 2009).

The first-order correction to $T_{S2w}^2$, that is, $T_{S2w}^2(1) = T_{S2w}^2/\bar{\tilde{\lambda}}$, has an asymptotic $\chi_2^2$ distribution under $H_0$, where $\bar{\tilde{\lambda}}$ is the average of the 2 nonzero eigenvalues of the matrix $\hat{\Sigma} I^{-1}(\hat{\theta}_{w0})$. A more accurate second-order correction to $T_{S2w}^2$, that is, $T_{S2w}^2(2) = T_{S2w}^2/\{\bar{\tilde{\lambda}}(1 + \hat{a}^2)\}$, has an asymptotic $\chi_{2/(1+\hat{a}^2)}^2$ distribution under $H_0$, where $\hat{a}^2 = (1/2) \sum_{i=1}^2 (\hat{\lambda}_i - \bar{\tilde{\lambda}})^2/\bar{\tilde{\lambda}}^2$ (Rao and Scott, 1984). A Satterthwaite $F$-version of $T_{S2w}^2$ is $FT_{S2w}^2(2) = T_{S2w}^2(2)/\{2/(1 + \hat{a}^2)\}$, which is asymptotically distributed as $F_{2/(1+\hat{a}^2), f}$ under $H_0$, where $f = \sum_{h=1}^H m_h - H$ (Thomas and Rao, 1987).

The quasi-efficient score tests require population parameters $f_0$ and $\mathbf{g}$. When established population parameters are available, they will be used. Otherwise, $f_0$ and $\mathbf{g}$ will be estimated from the survey data that are used for testing the genetic association.

## 3. SIMULATION STUDIES

Three intraclass correlation structures are evaluated in our simulation: (1) no intraclass correlation in case status and genotype, (2) intraclass correlation in case status but not in genotype, and (3) intraclass correlation in case status and genotype. We generated cases and controls that are correlated within each cluster with correlation coefficient of 0.09, and genotype data are generated with intraclass correlation coefficient of 0.017 assuming Hardy–Weinberg equilibrium for the whole population and within each cluster.

The Type I error rate and power of proposed test procedures are evaluated via Monte Carlo simulation with a significance level of 0.05. Details of simulation are described in the supplementary material available at *Biostatistics* online.

F-version test statistics that assume additive model or dominant model, that is, $FT_{Tw}^2(0.5)$ and $FT_{S1w}^2(0.5)$ or $FT_{Tw}^2(1)$ and $FT_{S1w}^2(1)$, maintain the nominal level for various parameter settings. When estimated $f_0$ and $\mathbf{g}$ are used instead of true $f_0$ and $\mathbf{g}$, quasi-efficient score tests achieve virtually the same Type I error rates with relative differences ranging from $<1\%$ to about 5%.

Compared to test statistics with varying sample weights, test statistics with common weights have more power, which would be expected since weighting tends to increase variances (Korn and Graubard, 1999, pp 172–177). Test statistics have more power when there is no intraclass correlation in case status and genotype compared to when case status and genotypes are correlated within each cluster. When a

genetic model is assumed, the tests have greater power under the correct model than under an incorrect model. When the data are generated using recessive or dominant models and incorrect genetic models are used for testing the association, tests based on additive model, that is, $T^2_{\text{Tw}}(0.5)$ test and $T^2_{\text{S1w}}(0.5)$ test and the corresponding F-version tests, are more powerful in general. All tests tend to have less power for small minor allele frequency and/or when there is intraclass correlation. Using estimated $f_0$ and **g** for quasi-efficient score tests has minimal impact on power compared with the true $f_0$ and **g**.

## 4. Real data analysis

The proposed methods are applied to NHANES III genetic data for 3 loci, that is, ADRB2 (rs1042713), TGFB1 (rs1982073), and VDR (rs2239185). Variances are estimated using pseudo-strata and pseudo-PSUs supplied by National Center for Health Statistics, Centers for Disease Control and Prevention (CDC).

There is disagreement in the literature whether or not there is an association between ADRB2 (rs1042713) and obesity. Some studies have shown that ADRB2 (rs1042713) is associated with obesity, while a recent study suggests that there is no association between ADRB2 (rs1042713) and obesity (see Jalba *and others*, 2008, for a meta-analysis). In our analysis, obesity is defined as BMI $\geqslant 30$, where BMI is body weight in kilogram divided by height in square meters. There are 6930 individuals with genotype information available for ADRB2. Among them, 1290 (18.61%) individuals have genotype $AA$, 3387 (48.87%) have genotype $AG$, and 2253 (32.51%) have genotype $GG$. About 1797 (25.15%) individuals have BMI $\geqslant 30$. The age-adjusted prevalence of obesity is about 23% (National Center for Health Statistics, 2008b).

TGFB1 (rs1982073) is shown to be associated with severe asthma (de Faria *and others*, 2008). Phase II of NHANES III has 6920 individuals with genotype information for TGFB1 (1452 $CC$, 3257 $CT$, and 2211 $TT$) and has 509 (7.11%) individuals with asthma, which is consistent with a national asthma prevalence of 7.2% estimated by National Center for Health Statistics (2008c).

VDR (rs2239185) has been found to modify lead toxic kinetics (Onalaja and Claudio, 2000) and may be associated with blood lead level (BLL). Our analysis has 6794 individuals with genotype information for VDR (1733 $CC$, 3245 $CT$, and 1816 $TT$) and 199 (2.82%) individuals with BLL $\geqslant 10$ µg/dl, which is consistent with CDC's estimate. According to CDC (1997), among those aged greater than or equal to 1 year, approximately 2.2% had BLLs greater than or equal to 10µg/dl (CDC, 1997).

Table 1 displays the $P$ values of the proposed tests for associations between ADRB2 and obesity, between TGFB1 and asthma, and between VDR and BLL. External $f'_0 s$ discussed above, which are consistent with the ones estimated from NHANES III, are used in our analysis. Since no external source of **g** was found, genotype frequencies **g** are estimated from the survey data. No significant associations were found between ADRB2 (rs1042713) and obesity, which is consistent with the finding of Jalba *and others* (2008), and between VDR (rs2239185) and high BLL. TGFB1 (rs1982073) was found to be associated with asthma at 5% level, which agrees with the finding of de Faria *and others* (2008).

## 5. Discussion

NHANES III provides us a unique opportunity to investigate the association between a wide variety of health factors and genetic variations at the US population level. In this paper, we examine trend test statistics with and without assuming a genetic model that are suitable for complex samples, which utilize multistage stratified cluster sample designs and sample weighting due to unequal probabilities of sample selection.

The Type I error rate and power of all test statistics considered allow for sample weighting and are evaluated via Monte Carlo simulations under noninformative sample weights with an alpha level of 0.05.

Table 1. *P values for association test for loci ADRB2, TGFB1, and VDR*

| | | ADRB2 | TGFB1 | VDR |
|---|---|---|---|---|
| Trend test (1 df) | $T^2_{\text{Tw}}(0)$ | 0.743 | 0.099 | 0.375 |
| | $T^2_{\text{Tw}}(0.5)$ | 0.613 | 0.012 | 0.827 |
| | $T^2_{\text{Tw}}(1)$ | 0.363 | 0.022 | 0.711 |
| | $FT^2_{\text{Tw}}(0)$ | 0.745 | 0.113 | 0.385 |
| | $FT^2_{\text{Tw}}(0.5)$ | 0.618 | 0.020 | 0.829 |
| | $FT^2_{\text{Tw}}(1)$ | 0.372 | 0.031 | 0.715 |
| Score test (1 df) | $T^2_{\text{S1w}}(0)$ | 0.744 | 0.102 | 0.372 |
| | $T^2_{\text{S1w}}(0.5)$ | 0.616 | 0.013 | 0.827 |
| | $T^2_{\text{S1w}}(1)$ | 0.368 | 0.023 | 0.710 |
| | $FT^2_{\text{S1w}}(0)$ | 0.746 | 0.115 | 0.381 |
| | $FT^2_{\text{S1w}}(0.5)$ | 0.621 | 0.021 | 0.829 |
| | $FT^2_{\text{S1w}}(1)$ | 0.377 | 0.032 | 0.714 |
| Score test (2 df) | $FT^2_{\text{S2w}}(2)$ | 0.480 | 0.046 | 0.593 |

Type I error rates are well controlled for F-version test statistics that assume an additive model or a dominant model. Test statistics with the correct genetic model achieve more power. When the genetic model is incorrect, $T^2_{\text{Tw}}(0.5)$ test, $T^2_{\text{S1w}}(0.5)$ test, and corresponding F-version tests achieve more power in general. When no genetic model is assumed, $FT^2_{\text{S2w}}(2)$ test maintains the nominal level excepted for small minor allele frequency. Quasi-score tests require population parameters $f_0$ and $\mathbf{g}$. Estimated $f_0$ and $\mathbf{g}$ have minimal impact on Type I error rates and power compared to true $f_0$ and $\mathbf{g}$. We recommend to use $FT^2_{\text{Tw}}(0.5)$ and $FT^2_{\text{S1w}}(0.5)$ to test the association between a disease and a candidate gene in complex sampling setting when genetic model is unknown and to use trend test statistic with correct genetic model specified when the genetic model is known. Since there is low power to determine if the sampling weighting is informative or not (Korn and Graubard, 1999), our recommended test statistics use the sample weights for testing for trend. There may be surveys that have highly inefficient weighting where modeling the sample selection would offer an alternative approach but this would require further research.

## SUPPLEMENTARY MATERIAL

Supplementary material is available at http://biostatistics.oxfordjournals.org.

## REFERENCES

ARMITAGE, P. (1955). Tests for linear trends in proportions and frequencies. *Biometrics* **11**, 375–386.

BRIER, S. S. (2007). Analysis of contingency tables under cluster sampling. *Biometrika* **67**, 591–596.

CDC (1997). *http://www.cdc.gov/mmwr/preview/mmwrhtml/00048339.htm.*

DE FARIA, I. C., DE FARIA, E. J., TORO, A. A., RIBEIRO, J. D. AND BERTUZZO, C. S. (2008). Association of TGF-beta1, CD14, IL-4, IL-4R and ADAM33 gene polymorphisms with asthma severity in children and adolescents. *Jornal de Pediatria (Rio J)* **84**, 203–210.

EPSTEIN, M. P., ALLENM, A. S. AND SATTEN, G. (2007). A simple and improved correction for population stratification in case-control studies. *American Journal of Human Genetics* **80**, 921–930.

EZZATI, T. M., MASSEY, J. T., WAKSBERG, J., CHU, A. AND MAURER, K. R. (1992). Sample design: Third National Health and Nutrition Examination Survey. *Vital and Health Statistics* **2**, 113.

FREIDLIN, B., ZHENG, G., LI, Z. AND GASTWIRTH, J. L. (2002). Trend tests for case-control studies of genetic markers: power, sample size and robustness. *Human Heredity* **53**, 146–152.

GRAUBARD, B. I. AND KORN, E. L. (1993). Hypothesis testing with complex survey data: the use of classical quadratic test statistics with particular reference to regression problems. *Journal of the American Statistical Association* **88**, 629–641.

JALBA, M. S., RHOADS, G. G. AND DEMISSIE, K. (2008). Association of codon 16 and codon 27 beta 2-adrenergic receptor gene polymorphisms with obesity: a meta-analysis. *Obesity* **16**, 2096–2106.

KORN, E. L. AND GRAUBARD, B. I. (1999). *Analysis of Health Surveys*. New York: John Wiley.

LI, Y. AND GRAUBARD, B. I. (2009). Testing Hardy-Weinberg equilibrium and homogeneity of Hardy-Weinberg disequilibrium using complex survey data. *Biometrics*. doi: 10.1111/j. 1541-0420.2009.01199.

LI, Z., GASTWIRTH, J. L. AND GAIL, M. H. (2005). Power and related statistical properties of conditional likelihood score tests for association studies in nuclear families with parental genotypes. *Annals of Human Genetics* **69**, 296–314.

MORTON, N. E. AND COLLINS, A. (1998). Tests and estimates of allelic association in complex inheritance. *Proceedings of the National Academy of Sciences USA* **95**, 11389–11393.

NATIONAL CENTER FOR HEALTH STATISTICS (2008a). *http://www.cdc.gov/nchs/about/major/nhanes/nh3data_genetic.htm.*

NATIONAL CENTER FOR HEALTH STATISTICS (2008b). *http://www.cdc.gov/nchs/products/pubs/pubd/hestats/obese/obse99.htm.*

NATIONAL CENTER FOR HEALTH STATISTICS (2008c). *http://www.cdc.gov/nchs/products/pubs/pubd/hestats/asthma/asthma.htm.*

ONALAJA, A. O. AND CLAUDIO, L. (2000). Genetic susceptibility to lead poisoning. *Environmental Health Perspectives* **108**(Suppl 1), 23–28.

RAO, J. N. K. AND SCOTT, A. J. (1984). On chi-square tests for multiway contingency tables with cell proportions estimated from survey data. *Annals of Statistics* **12**, 46–60.

RAO, J. N. K., SCOTT, A. J. AND SKINNER, C. J. (1998). Quasi-score tests with survey data. *Statistica Sinica* **8**, 1059–1070.

RYCKMAN, K. K., JIANG, L., LI, C., BARTLETT J., HAINES, J. L. AND WILLIAMS, S. M. (2008). A prevalence-based association test for case-control studies. *Genetic Epidemiology* **32**, 600–605.

SASIENI, P. D. (1997). From genotypes to genes: doubling the sample size. *Biometrics* **53**, 1253–1261.

SHAH, B. V. (2004). Comment on 'Linearization variance estimators for survey data' by A. Demnati and J. N. K. Rao. *Survey Methodology* **30**, 16.

SLAGER, S. L. AND SCHAID, D. J. (2001). Case-control studies of genetic markers: power and sample size approx-imations for Armitage's test for trend. *Human Heredity* **52**, 149–153.

THOMAS, D. R. AND RAO, J. N. K. (1987). Small-sample comparisons of level and power for simple goodness-of-fit statistics under cluster sampling. *Journal of the American Statistical Association* **82**, 630–636.

ZHENG, G. AND GASTWIRTH, J. L. (2006). On estimation of the variance in Cochran-Armitage trend tests for genetic association using case-control studies. *Statistics in Medicine* **25**, 3150–3159.