

Note

Group Sequential Methods and Sample Size Savings in Biomarker-Disease Association Studies

R. Aplenc,^{*,†,1} H. Zhao,[†] T. R. Rebbeck^{*} and K. J. Propert^{*}

^{*}Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104-6021 and [†]Children's Hospital of Philadelphia, Philadelphia, Pennsylvania 19104

Manuscript received September 27, 2002

Accepted for publication December 9, 2002

ABSTRACT

Molecular epidemiological association studies use valuable biosamples and incur costs. Statistical methods for early genotyping termination may conserve biosamples and costs. Group sequential methods (GSM) allow early termination of studies on the basis of interim comparisons. Simulation studies evaluated the application of GSM using data from a case-control study of GST genotypes and prostate cancer. Group sequential boundaries (GSB) were defined in the EAST-2000 software and were evaluated for study termination when early evidence suggested that the null hypothesis of no association between genotype and disease was unlikely to be rejected. Early termination of *GSTM1* genotyping, which demonstrated no association with prostate cancer, occurred in >90% of the simulated studies. On average, 36.4% of biosamples were saved from unnecessary genotyping. In contrast, for *GSTT1*, which demonstrated a positive association, inappropriate termination occurred in only 6.6%. GSM may provide significant cost and sample savings in molecular epidemiology studies.

ALTHOUGH group sequential methods (GSM) are routinely used to monitor randomized clinical trials, they have not yet been widely applied to molecular epidemiology (ME) studies. In clinical trials, GSM allow early closure of one or more treatment arms on the basis of interim analysis (WHITEHEAD 1999). By enabling early closure, GSM protect patients from unnecessary exposure to an unfavorable treatment. The statistical "cost" for early closure, the loss of precision of the effect size estimate, is acceptable because patients are protected from unnecessary exposure to unfavorable therapies. There is an extensive literature on group sequential methods and their application to clinical trials; an excellent summary is provided in JENNISON and TURNBULL (2000).

Early closure for "futility," in which the study is unlikely to lead to rejection of the null hypothesis, is becoming more commonly used in clinical trials. Although ME studies lack this ethical imperative for early closure, such studies would benefit from early closure for futility for several reasons. First, such studies often use biologic samples that are difficult to obtain or limited in quantity. Second, genotype assessment incurs both material and labor costs. Thus, early closure for failure to reject the

null hypothesis may save samples, reagents, labor, and opportunity costs. Finally, clearly defined interim analysis procedures would provide investigators with a formal tool for evaluating their data on an ongoing basis.

Previous investigators have described the importance of early closure for null effects (GOULD 1983; JENNISON and TURNBULL 2000). O'Neill and Anello first described the use of the Wald sequential probability ratio test (SPRT), an open procedure, in a matched case-control study (O'NEILL and ANELLO 1978). PASTERNAK and SHORE (1980) demonstrated that in a cohort study the group sequential design had generally higher efficiency than that of a fixed sample plan. KAAKS *et al.* (1994) demonstrated the application of a sequential *t*-test to the use of biologic samples. VAN DER TWEEL and VAN NOORD (2000) described both a SPRT and a triangular test for sample sequential analysis of genotype data. Recently, SATAGOPAN *et al.* (2002) described the use of a two-stage design for maximizing power when the total cost is the primary study constraint.

Current molecular epidemiology studies, however, have practical characteristics that preclude these approaches. First, the finite number of available samples and limits on funding time lines prevent the use of an "open" GSM whose sample size is potentially unlimited. Second, almost all molecular epidemiology studies acquire genotype data on a group of samples simultaneously. Thus, the most appropriate GSM must evaluate

¹Corresponding author: Center for Clinical Epidemiology and Biostatistics, 9th Fl., Blockley Hall, 423 Guardian Dr., Philadelphia, PA 19104-6021. E-mail: raplenc@ceeb.med.upenn.edu

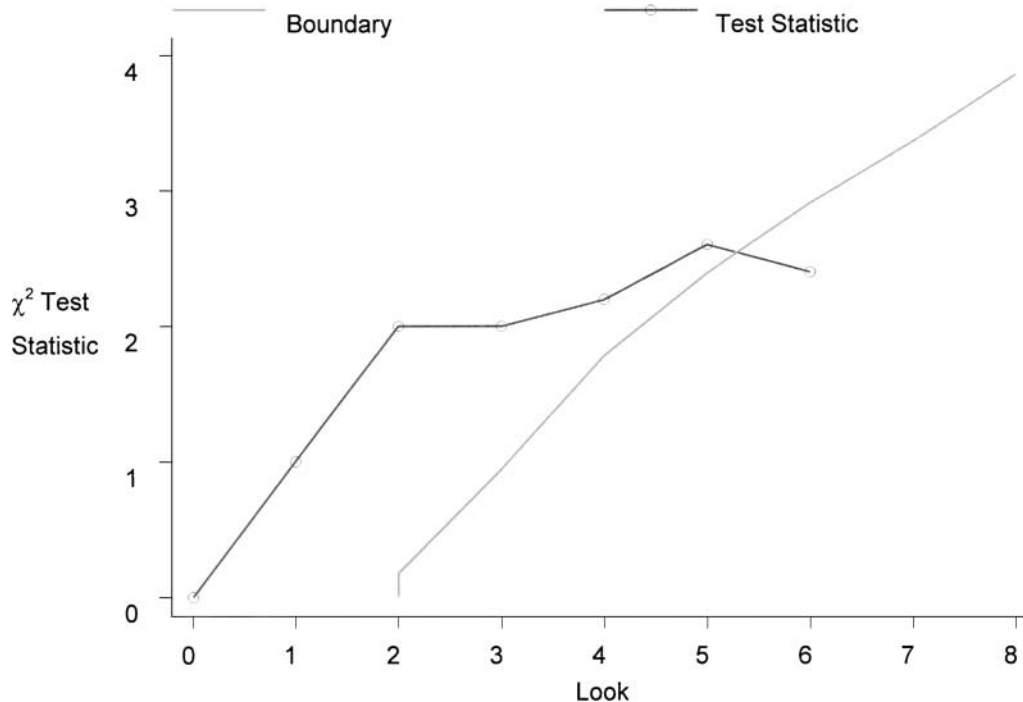


FIGURE 1.—Hypothetical OBF boundary and test statistic. Evolution of two-sided test statistic that crosses the OBF boundary at the sixth look and terminates the experiment.

sequential groups of genotype data rather than sequential individual genotypes. Finally, current studies often evaluate a small number of genotypes (<10), thus making the sample itself the primary limiting variable.

We evaluated the group sequential boundaries methods because of their widespread use and the availability of GSM commercial software. In GSM, the number of interim “looks” is frequently equally spaced and predefined at the design stage. These criteria may be relaxed during study conduct. In a case-control study, the test statistic is the χ^2 value corresponding to the odds ratio of disease between cases and controls. In the case of early stopping for futility, if the χ^2 test statistic is less than a predefined value, called a boundary value, then it is unlikely that genotyping additional samples will give a statistically significant result. Therefore genotyping stops once the χ^2 test statistic crosses this boundary. Stopping boundaries may be defined by commercial software packages such as EAST-2000 (Cambridge, MA; <http://www.cytel.com>) or PEST (Reading, UK; http://www.rdg.ac.uk/mps/mps_home/software/pest4/pest4.htm) or by writing local software (SCHOENFELD 2001).

Figure 1 demonstrates the evolution of a test statistic in a hypothetical study with eight looks. The study would terminate early to accept the null hypothesis if the path of the test statistic crossed the boundary at any point, as occurs at look number 6. For some choices of parameter values, early closure is not possible. For example, the boundary shown in Figure 1 does not allow closure at the first look, where it is undefined. Therefore, irrespective of the results obtained at the first look, a second round of genotyping would be required.

Simulation studies were used to evaluate the applica-

tion of GSM. Two previously published data sets of GST genotype and prostate cancer risk were used for the simulations (REBBECK *et al.* 1999). These data sets were chosen for several reasons. First, the *GSTMI* data set reported a null association and represented the case where early stopping for futility with a GSM could provide significant sample and cost savings. Second, the *GSTTI* data set reported a positive association and was used to evaluate the frequency of inappropriate genotyping termination. Since publication, additional cases and controls have been genotyped; the sample set used in the simulations contained a total of 675 *GSTMI* and 725 *GSTTI* genotypes. The observed odds ratio (OR) for *GSTMI* was OR = 0.99, 95% confidence interval (C.I.) 0.72–1.38; for *GSTTI*, the OR = 1.61, 95% C.I. 1.12–2.32. In addition to representing both a null and a positive association, both data sets have sample sizes and odds ratios typically seen in present-day ME studies. Finally, the raw data were readily available for the simulation studies.

O’Brien-Fleming (OBF) stopping boundaries for both rejection and failure of rejection of the null hypothesis at each interval of genotype data acquisition were defined using EAST-2000 (O’BRIEN and FLEMING 1979). We chose the OBF boundary because it is most frequently used to monitor clinical trials and is more conservative than the alternative Pocock boundary for both rejection and acceptance of the null hypothesis. A more conservative boundary also limits the decrease in power associated with using GSM. We chose EAST-2000 for its ease of use and commercial availability since many groups conducting molecular epidemiology studies did not have resources to generate in-house software.

TABLE 1
Results of simulation studies for *GSTMI*

| No. | Simulation parameters | | | | | No. of replications stopping at each interim look | | | | | | Simulation results | | Fixed design (<i>N</i>) |
|-----|-----------------------|------|----------|-----|-------|---|-----|-----|-----|-----|-----|--------------------|------------------|---------------------------|
| | Control | Case | Look no. | OR | Power | 1 | 2 | 3 | 4 | 5 | 6 | % stop early | Average <i>N</i> | |
| 1 | 0.10 | 0.15 | 2 | 1.6 | 46.1 | 947 | | | | | | 94.7 | 377 | 1366 |
| 2 | 0.10 | 0.17 | 2 | 1.8 | 69.0 | 855 | | | | | | 85.5 | 406 | 741 |
| 3 | 0.10 | 0.18 | 2 | 2 | 85.2 | 684 | | | | | | 68.4 | 459 | 583 |
| 4 | 0.38 | 0.50 | 2 | 1.6 | 84.3 | 708 | | | | | | 70.8 | 452 | 530 |
| 5 | 0.38 | 0.53 | 2 | 1.8 | 96.6 | 239 | | | | | | 23.9 | 600 | 339 |
| 6 | 0.38 | 0.55 | 2 | 2 | 99.4 | 45 | | | | | | 4.5 | 661 | 263 |
| 7 | 0.50 | 0.62 | 2 | 1.6 | 84.4 | 708 | | | | | | 70.8 | 452 | 530 |
| 8 | 0.50 | 0.64 | 2 | 1.8 | 96.5 | 248 | | | | | | 24.8 | 597 | 385 |
| 9 | 0.50 | 0.67 | 2 | 2 | 99.4 | 50 | | | | | | 5.0 | 659 | 256 |
| 10 | 0.10 | 0.15 | 3 | 1.6 | 42.9 | 821 | 179 | | | | | 100.0 | 318 | 1366 |
| 11 | 0.10 | 0.17 | 3 | 1.8 | 67 | 508 | 492 | | | | | 100.0 | 403 | 741 |
| 12 | 0.10 | 0.18 | 3 | 2 | 84.7 | 101 | 898 | | | | | 99.9 | 513 | 583 |
| 13 | 0.38 | 0.50 | 3 | 1.6 | 83.7 | 128 | 871 | | | | | 99.9 | 506 | 530 |
| 14 | 0.38 | 0.53 | 3 | 1.8 | 96.3 | NA | 999 | | | | | 99.9 | 540 | 339 |
| 15 | 0.38 | 0.55 | 3 | 2 | 99.4 | NA | 996 | | | | | 99.6 | 541 | 263 |
| 16 | 0.50 | 0.62 | 3 | 1.6 | 83.8 | 124 | 875 | | | | | 99.9 | 507 | 530 |
| 17 | 0.50 | 0.64 | 3 | 1.8 | 96.1 | NA | 999 | | | | | 99.9 | 540 | 385 |
| 18 | 0.50 | 0.67 | 3 | 2 | 99.3 | NA | 996 | | | | | 99.6 | 541 | 256 |
| 19 | 0.10 | 0.15 | 7 | 1.6 | 38.3 | 104 | 416 | 309 | 124 | 37 | 9 | 99.9 | 235 | 1366 |
| 20 | 0.10 | 0.17 | 7 | 1.8 | 63.6 | NA | 96 | 452 | 327 | 105 | 19 | 99.9 | 315 | 741 |
| 21 | 0.10 | 0.18 | 7 | 2 | 82 | NA | NA | 180 | 545 | 230 | 44 | 99.9 | 373 | 583 |
| 22 | 0.38 | 0.50 | 7 | 1.6 | 80.9 | NA | NA | 218 | 518 | 222 | 41 | 99.9 | 368 | 530 |
| 23 | 0.38 | 0.53 | 7 | 1.8 | 95.6 | NA | NA | NA | 345 | 542 | 112 | 99.9 | 429 | 339 |
| 24 | 0.38 | 0.55 | 7 | 2 | 99.3 | NA | NA | NA | 74 | 671 | 252 | 99.7 | 467 | 263 |
| 25 | 0.50 | 0.62 | 7 | 1.6 | 81 | NA | NA | 213 | 521 | 224 | 41 | 99.9 | 369 | 530 |
| 26 | 0.50 | 0.64 | 7 | 1.8 | 95.4 | NA | NA | NA | 359 | 531 | 109 | 99.9 | 428 | 385 |
| 27 | 0.50 | 0.67 | 7 | 2 | 99.1 | NA | NA | NA | 86 | 671 | 240 | 99.7 | 465 | 256 |

The sample pool for *GSTMI* simulations was 675. Control indicates the genotype frequency in controls; Case, genotype frequency in cases; Look no., number of looks in each simulation; OR, odds ratio defined by genotype frequencies; Power, simulation power; % stop early, percentage of simulations that terminated prior to genotyping all *N* samples for a given set of simulation parameters; Average *N*, the average number of samples genotyped at simulation termination over 1000 replications; *N*, number of samples necessary for 80% power at the specified control genotype frequencies and OR; NA, not applicable, as early termination not allowed at this look.

Although EAST-2000 requires that boundaries for rejection of the null hypothesis be part of the overall calculation of GSM boundaries, only boundaries for failure of rejection of null hypothesis were used in the simulations.

For all simulations, the overall two-sided type I error was set at $\alpha = 0.05$. Since the sample pool was fixed ($N = 675$ for *GSTMI* and 725 for *GSTTI*), the power was defined by the sample size, null genotype frequencies in controls, and OR. We chose not to specify a type II error rate to examine the performance of the GSM method over a range of genotype frequencies and ORs. Genotype frequencies in controls were set at 10%, at 50%, and at the genotype frequency observed in the data set used. The observed genotype frequencies were 38% for *GSTMI* and 28% for *GSTTI*. ORs of 1.6, 1.8, and 2.0 were examined. The OR of 1.6 was chosen to correspond to that observed for *GSTTI*. An OR of 2.0 corre-

sponds to that often used as the target “clinically significant association” for many epidemiological studies. The OR of 1.8 was chosen to be intermediate between these two. For these simulations, the interval of genotype data acquisition was termed a “look.” Each look contained a multiple of 90 genotypes to simulate genotype acquisition from a 96-well PCR-based genotyping method (*e.g.*, 90 genotypes and 6 control samples per PCR run).

In addition to the simulation parameters defined by the baseline frequencies and OR, three different look strategies were examined. The first strategy had two looks, with the interim look occurring after ~50% of the samples had been genotyped. The second strategy used the maximum number of possible looks, given the sample size and the restriction that each look (except the last) must include a multiple of 90 samples. The third strategy chosen was intermediate between these. Thus simulations for *GSTMI* examined two, three, or

TABLE 2
Results of simulation studies for *GSTTI*

| No. | Simulation parameters | | | | | No. of replications stopping at each interim look | | | | | | | Simulation results | | Fixed design (<i>N</i>) |
|-----|-----------------------|-------|----------|-----|-------|---|-----|----|----|----|----|---|--------------------|------------------|---------------------------|
| | Controls | Cases | Look no. | OR | Power | 1 | 2 | 3 | 4 | 5 | 6 | 7 | % stop early | Average <i>N</i> | |
| 1 | 0.10 | 0.15 | 2 | 1.6 | 44.8 | 156 | | | | | | | 15.6 | 668 | 1514 |
| 2 | 0.10 | 0.17 | 2 | 1.8 | 68.4 | 56 | | | | | | | 5.6 | 705 | 804 |
| 3 | 0.10 | 0.18 | 2 | 2 | 85.3 | 21 | | | | | | | 2.1 | 718 | 627 |
| 4 | 0.28 | 0.38 | 2 | 1.6 | 78.1 | 37 | | | | | | | 3.7 | 712 | 623 |
| 5 | 0.28 | 0.41 | 2 | 1.8 | 93.5 | 7 | | | | | | | 0.7 | 722 | 399 |
| 6 | 0.28 | 0.44 | 2 | 2 | 98.7 | 0 | | | | | | | 0.0 | 725 | 310 |
| 7 | 0.50 | 0.62 | 2 | 1.6 | 79.6 | 31 | | | | | | | 3.1 | 714 | 638 |
| 8 | 0.50 | 0.64 | 2 | 1.8 | 94.2 | 6 | | | | | | | 0.6 | 723 | 466 |
| 9 | 0.50 | 0.67 | 2 | 2 | 98.7 | 0 | | | | | | | 0.0 | 725 | 313 |
| 10 | 0.10 | 0.15 | 4 | 1.6 | 39.6 | 149 | 104 | 18 | | | | | 27.1 | 603 | 1514 |
| 11 | 0.10 | 0.17 | 4 | 1.8 | 65.9 | 13 | 71 | 23 | | | | | 10.7 | 688 | 804 |
| 12 | 0.10 | 0.18 | 4 | 2 | 83.6 | NA | 27 | 23 | | | | | 5.0 | 711 | 627 |
| 13 | 0.28 | 0.38 | 4 | 1.6 | 76 | NA | 44 | 27 | | | | | 7.1 | 704 | 623 |
| 14 | 0.28 | 0.41 | 4 | 1.8 | 92.6 | NA | 8 | 11 | | | | | 1.9 | 720 | 399 |
| 15 | 0.28 | 0.44 | 4 | 2 | 98.4 | NA | 0 | 4 | | | | | 0.4 | 724 | 310 |
| 16 | 0.50 | 0.62 | 4 | 1.6 | 77.6 | NA | 41 | 24 | | | | | 6.5 | 706 | 638 |
| 17 | 0.50 | 0.64 | 4 | 1.8 | 93.4 | NA | 7 | 11 | | | | | 1.8 | 720 | 466 |
| 18 | 0.50 | 0.67 | 4 | 2 | 98.4 | NA | NA | 4 | | | | | 0.4 | 724 | 313 |
| 19 | 0.10 | 0.15 | 8 | 1.6 | 36.6 | 52 | 135 | 79 | 33 | 27 | 9 | 0 | 33.5 | 561 | 1514 |
| 20 | 0.10 | 0.17 | 8 | 1.8 | 62.6 | NA | 16 | 60 | 34 | 28 | 16 | 0 | 15.4 | 666 | 804 |
| 21 | 0.10 | 0.18 | 8 | 2 | 81.8 | NA | NA | 17 | 20 | 20 | 15 | 1 | 7.3 | 702 | 627 |
| 22 | 0.28 | 0.38 | 8 | 1.6 | 73.4 | NA | NA | 38 | 26 | 29 | 19 | 0 | 11.2 | 687 | 623 |
| 23 | 0.28 | 0.41 | 8 | 1.8 | 91.6 | NA | NA | NA | 8 | 16 | 11 | 0 | 3.5 | 716 | 399 |
| 24 | 0.28 | 0.44 | 8 | 2 | 98.2 | NA | NA | NA | NA | 6 | 5 | 1 | 1.2 | 722 | 310 |
| 25 | 0.50 | 0.62 | 8 | 1.6 | 75.1 | NA | NA | 30 | 22 | 25 | 18 | 1 | 9.6 | 693 | 638 |
| 26 | 0.50 | 0.64 | 8 | 1.8 | 92.5 | NA | NA | NA | 6 | 13 | 12 | 1 | 3.2 | 717 | 466 |
| 27 | 0.50 | 0.67 | 8 | 2 | 98.1 | NA | NA | NA | NA | 6 | 5 | 1 | 1.2 | 722 | 313 |

The sample pool for *GSTTI* simulations was 725. Control indicates the genotype frequency in controls; Case, genotype frequency in cases; Look no., number of looks in the simulation; OR, odds ratio defined by genotype frequencies; Power, simulation power; % stop early, percentage of simulations that terminated prior to genotyping all *N* samples for a given set of simulation parameters; Average *N*, the average number of samples genotyped at simulation termination over 1000 replications; *N*, number of samples necessary for 80% power at the specified control genotype frequencies and OR; NA, not applicable, as early termination not allowed at this look.

seven looks; two, four, or eight looks were examined for *GSTTI*.

A total of 1000 replications were performed for each of the 27 combinations of baseline gene frequency, OR, and number of looks. Simulations were done separately for the *GSTMI* and *GSTTI* data sets. For each replication, prostate cancer cases and controls were randomly sampled from the true data sets without replacement and in proportion to their relative frequencies. The observed OR and χ^2 test statistic were calculated for each look. The χ^2 test statistic was then compared to the boundary value calculated by EAST-2000 for study termination. If the test statistic was less than the boundary for early stopping, *i.e.*, if the test statistic “crossed the boundary,” then the run terminated. If the test statistic did not cross the boundary, then an additional look was selected and the test statistic recalculated, accounting for the information gained in the prior look. This

procedure was repeated until the test statistic crossed a boundary or all genotypes were sampled. All simulation studies and analyses were performed using STATA v7.0 (College Park, TX).

In the above, we dealt with the potential for early closure by using the boundary values themselves (on the χ^2 test statistic scale). This method allows application of these methods to test statistics that are not built into standard group sequential software packages. However, it should be noted that an alternative means of conducting monitoring of a molecular epidemiology trial would be to use directly the methods developed for a comparison of two binomials. These methods are available in, for example, EAST-2000.

Results for *GSTMI* simulations are shown in Table 1. Overall, 91.5% of the simulations terminated early with a range of 4.5–100%. The median genotyped sample size was 459. Thus, use of GSM decreased the median

sample size by 32%. Results for the *GSTTI* simulations are shown in Table 2. On average, only 6.6% of the *GSTTI* simulations terminated early. The median sample size was 714 with the sample size of 725 representing the entire data set. This low frequency of termination is appropriate as an association between *GSTTI* genotype and outcome was present in the data set.

Our simulations indicate that GSM may provide significant improvements for case-control molecular epidemiology studies. Our approach of evaluating genotype data in multiples of 90 more closely reflects laboratory data acquisition and is thus directly applicable to large molecular epidemiology studies. For *GSTTI* simulations with 80% power, assuming a genotype cost of \$3.00/genotype, the use of GSM would save ~\$650 from a total cost of \$2025, in addition to savings in technician time and reagents. This sample size savings had a relatively small cost to the overall power of the study. The average difference in study power between a fixed sample design and a GSM design for *GSTMI* simulations with 80% power was 3.3% (average fixed sample size power was 86.2%; average GSM design power was 82.9%). For these simulations, the average difference in study power between a one-look and a maximum-look strategy was also small—3.3%.

A number of observations may be made regarding the effects of varying model parameters on the probability of early stopping. First, the frequency of early stopping decreased as the study power increased. Although power is affected by the baseline frequency, OR, and sample size, the frequency of early stopping was “monotonic” in power. Thus, in all cases lower-power studies had higher rates of termination and terminated at earlier looks than did models with higher power. This corresponds to the intuition that studies with low power should be more likely to close early because the *a priori* chance of finding a significant association is very small, even if an association exists. However, appropriately powered models closed appropriately early in the *GSTMI* simulations and had low rates of inappropriate closure in the *GSTTI* simulations.

The baseline genotype frequency in controls (p_1) directly affects the statistical power. *GSTMI* models with a baseline frequency $p_1 = 0.38$ or $p_1 = 0.50$ and *GSTTI* models with $p_1 = 0.28$ had the highest power for a given OR and number of looks. These higher-power models closed later and had larger average sample numbers. Likewise, simulations with larger OR closed later and had larger average sample numbers than simulations with lower OR for the same p_1 and number of looks.

Finally, increasing the number of looks decreased the study power and in general decreased the average sample number. Interestingly, for *GSTMI* models with a typical power of ~80%, an intermediate number of looks had higher average sample numbers than models with either the minimum or the maximum number of

looks. Models with two looks obtained enough genotype information at the first look to close early with a high rate with attendant sample size savings. This is consistent with the results of similar analyses in clinical trials (POCOCK 1982). Models with seven looks gained the majority of genotype information in the middle looks, also allowing for substantial sample size savings. However, models with three looks had low rates of closure at look 1, thus requiring a second look.

Since our simulations indicate that an intermediate look strategy may give a higher average sample number for studies with ~80% power, investigators may wish to choose either a minimum- or a maximum-look strategy. Since the power cost of additional looks is relatively small, the optimal number of looks will be determined largely by the opportunity cost of multiple data analyses as well as by the need to conserve samples and costs. If samples are limited or expensive to assay, investigators may wish to perform multiple looks to minimize the average sample number. However, if sample conservation or cost minimization are not overriding concerns, then investigators may wish to perform only one interim analysis.

Supported by the Doris Duke Charitable Foundation (R.A.) and the Leonard and Madilyn Abramson Endowed Chair, National Institutes of Health grant R01-CA85074 (T.R.R.)

LITERATURE CITED

- GOULD, A. L., 1983 Abandoning lost causes (early termination of unproductive clinical trials). Proceedings of the Biopharmaceutical Sciences, American Statistical Association, Washington, DC, pp. 31–34.
- JENNISON, C., and B. W. TURNBULL, 2000 *Group Sequential Methods With Application to Clinical Trials*. Chapman Hall/CRC Press, New York.
- KAAS, R., I. VAN DER TWEEL, P. A. VAN NOORD and E. RIBOLI, 1994 Efficient use of biological banks for biochemical epidemiology: exploratory hypothesis testing by means of a sequential t-test. *Epidemiology* **5**: 429–438.
- O'BRIEN, P. C., and T. R. FLEMING, 1979 A multiple testing procedure for clinical trials. *Biometrics* **35**: 549–556.
- O'NEILL, R. T., and C. ANELLO, 1978 Case-control studies: a sequential approach. *Am. J. Epidemiol.* **108**: 415–424.
- PASTERNAK, B. S., and R. E. SHORE, 1980 Group sequential methods for cohort and case-control studies. *J. Chronic Dis.* **33**: 365–373.
- POCOCK, S. J., 1982 Interim analyses for randomized clinical trials: the group sequential approach. *Biometrics* **38**: 153–162.
- REBBECK, T. R., A. H. WALKER, J. M. JAFFE, D. L. WHITE, A. J. WEIN *et al.*, 1999 Glutathione S-transferase-mu (GSTM1) and -theta (GSTT1) genotypes in the etiology of prostate cancer. *Cancer Epidemiol. Biomarkers Prev.* **8**: 283–287.
- SATAGOPAN, J. M., D. A. VERBEL, E. S. VENKATRAMAN, K. E. OFFIT and C. B. BEGG, 2002 Two-stage designs for gene-disease association studies. *Biometrics* **58**: 163–170.
- SCHOENFELD, D. A., 2001 A simple algorithm for designing group sequential clinical trials. *Biometrics* **57**: 972–974.
- VAN DER TWEEL, I., and P. A. VAN NOORD, 2000 Sequential analysis of matched dichotomous data from prospective case-control studies. *Stat. Med.* **19**: 3449–3464.
- WHITEHEAD, J., 1999 A unified theory for sequential clinical trials. *Stat. Med.* **18**: 2271–2286.

