Practical and analytic aspects of using friend controls in case-control studies: experience from a case-control study of childhood cancer

Greta R Bunin[1], **Saran Vardhanabhuti**[2], **Agueda Lin**[1], **Greta L Anschuetz**[3], and **Nandita Mitra**[2]

[1]Division of Oncology and Center for Childhood Cancer Research, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania

[2]Department of Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, Pennsylvania, USA

[3]Division of Disease Control, Philadelphia Department of Public Health, Philadelphia, Pennsylvania

Summary

The authors report empirical data on the use of friend controls, specifically response rates, case-control concordance, and analytic approaches. The data derive from a North American multi-institutional study of childhood cancer that was conducted in 2002–2007 and that focused on paternal exposures. Case parents nominated friends as potential controls; up to 3 controls participated per case. For 137 (69%) of the 199 case families, at least 1 control participated. Of 374 potential controls contacted, 247 (66%) participated. Case fathers with controls were markedly more likely to be non-Hispanic white, college graduates, and non-smokers compared to case fathers without controls. Odds ratios adjusted for demographic characteristics were generally similar but occasionally differed between analyses that included only members of matched sets and those that included all participants, i.e. controls and cases with and without controls. For demographic characteristics, simulations demonstrated that the observed concordance of cases and controls within matched sets exceeded that expected under random ascertainment, indicating probable overmatching. However, the observed concordance of smoking and other exposures was similar to the expectation under random ascertainment suggesting little overmatching on exposures. Although not ideal, friend controls were convenient, had a reasonably high response rate, and provided controls closely matched on race/ethnicity, education, and age.

Friend controls are convenient as researchers need expend little effort to identify them through the cases.[1] Friends are presumed to be more motivated and have higher response rates compared to general population controls who do not know the case.[1] However, selection bias may occur because, if we think of individuals as having rosters of friends, more sociable individuals will appear on more rosters and introverted individuals on fewer or even no such rosters.[1] Thus, extroverts are more likely, and introverts less likely, to become friend controls. To the extent that the exposures of interest are associated with sociability or personality, selection bias that affects the results will occur.

Friend controls are, by definition, individually matched to cases, even when the choice of control group is made primarily for convenience and not purposefully as a method for

Corresponding author: Dr. Greta R Bunin, Children's Hospital of Philadelphia, 3535 Market St., Room 1472, Philadelphia, PA 19104; 215 590-1445 (voice), 215 590-3020 (fax); bunin@email.chop.edu.

controlling for likely confounders.[2] In other words, cases and controls may be inadvertently matched on characteristics not intended as matching factors. For example, friend controls may be matched on demographic factors such as socioeconomic status (SES) and on exposures such as smoking, drinking, and occupational substances, as these may be the basis of friendship or activities done as friends.[1, 3] Both intentional and unintentional matching factors may be correlated with exposure but not disease, i.e., the cases and controls may be matched on a factor that is not a confounder. If a matching factor is strongly associated with exposure, many case-control sets may be concordant for the exposure.[1, 4] This overmatching reduces power because concordant sets do not contribute to the estimation of risk ratios. Moreover, when cases and controls are closely matched, whether intentionally or unintentionally, the effect of the matching factors on disease cannot be examined.

In recent decades, epidemiologists have not considered friend controls as first choice controls because of their limitations. However, as approaches to obtaining general population controls such as random digit dialing (RDD) have become less feasible in recent years,[5] we have anecdotally observed increased interest in using friend controls, particularly for studies investigating the role of genetic polymorphisms. In addition, the use of friend controls may be particularly appealing for studies of parents' exposures in relation to childhood disease because some other possible sources of controls are less feasible for studying children compared to adults. For example, the use of population registries or driver's license files are less efficient in identifying and recruiting parents, as many non-parents have to be contacted in the process. A number of recent case-control studies of fertility, cancer, Alzheimer's disease, menstrual disturbances, epilepsy in children, and other conditions, many of which focused on genetic polymorphisms, have used friend controls.[6–23] The list of conditions demonstrates that friend controls have been used for all age groups, from children to the elderly.

Although the theoretical advantages and limitations of friend controls are well described, few publications report empirical data on proportion of cases for whom a friend control can be recruited, response rates, degree of concordance, and practical analytic issues. We report here on these topics based on our experience in a recent study of retinoblastoma, a cancer of young children.

METHODS

Study population

The institutional review boards (IRBs) of all institutions that recruited participants approved the study and all participants gave informed consent.

Children with retinoblastoma, a rare cancer of the retina occurring in young children, who were diagnosed from January 1998 – May 2006 were identified through 9 institutions in the US and Canada. We restricted the cases to those having the sporadic bilateral form of the disease, which develops as a result of a new mutation in the *RB1* gene that occurs before the child's conception. Although the new mutation may occur in either the sperm or egg, it most often occurs in the sperm.[24] Based on these facts, we hypothesized that parental exposures, particularly those of the father, prior to the child's conception increase the risk of the child developing retinoblastoma.

In addition to having a diagnosis of sporadic bilateral retinoblastoma, eligible cases lived in North America, had at least one parent who spoke English or Spanish, had at least one biological parent available (i.e., child not adopted or in foster care), and had at least one parent with phone service. Case mothers and fathers were interviewed by telephone in 2002 – 2007 about a variety of exposures and demographic characteristics including self-

identified race (using the standard US Census categories: white, African-American, Asian, American Indian/Alaska native, native Hawaiian/other Pacific Islander) and ethnicity (Hispanic/Latino vs. non-Hispanic/Latino). Race/ethnicity as well as education, income, and age were included as possible confounders of the relationship between exposure and disease. During the interview, the parent was asked to enumerate and provide the age of potentially eligible friends and cousins or other blood relatives of the case who were under the age of 15.

After the interview, study personnel chose up to 3 friends and up to 3 relatives who were in the same or adjacent age group (0–1, 2–4, 5–6, 7–9, 10–12 years) as the case and met the same eligibility criteria as the cases except for diagnosis of retinoblastoma. For relatives, we selected those families in which the father of the control was not a blood relative of the case father. Thus, the control fathers, whether of friends or relatives, were not biologically related to the case fathers and can be considered friend controls as described in the literature. If there were more than 3 friends or more than 3 relatives, 3 of each were chosen randomly. The interviewer re-contacted the parents and requested that they ask the selected friends and relatives if study personnel could contact them about the study. The interviewer contacted the parents again to obtain names, addresses, and phone numbers of the friends and relatives who agreed to be contacted about the study. We aimed to enroll 1 relative and 1 – 2 friends per case. If there was more than 1 potential control in a category (relative or friend), we randomly selected one to be contacted first. If the first potential control did not participate, we randomly selected another for contact. This process continued until a control was interviewed or contact had been attempted for all potential controls. For the remainder of this report, we will refer to all the controls as friend controls.

When contact information was obtained, study personnel mailed introductory letters to the potential control parents and followed up with a phone call during which the interviewer confirmed eligibility. An interviewer administered the risk factor questionnaire by phone to consenting parents of cases and controls. When a parent was not available, the other parent was asked to act as a proxy and complete the interview. Parents were offered $25 as reimbursement for their time and effort.

Modeling Approaches

If epidemiologists were asked to analyze data from a study with matched friend controls, we conjectured that most would use one of 3 approaches for estimating odds ratios (ORs) and 95% confidence intervals (CIs) for the relationships between the disease and selected exposures and characteristics. The first approach (referred to as 'matched') is a matched analysis of the case-control sets using conditional logistic regression. This approach discards all cases that do not have a control. The second approach (referred to as 'restricted unmatched') is an analysis that again only uses cases that have matched controls but models the data without consideration of matching by using unconditional logistic regression. The third analysis which we refer to as 'unrestricted unmatched' includes all participants, i.e. controls and all cases regardless of whether they had any controls, and uses unconditional logistic regression to model the data. We used all 3 approaches and compared the results.

Data analysis

The analysis reported here focused on fathers because the study focused on paternal exposures. Cases who had an interviewed control and those who did not were compared on demographic characteristics and selected exposures using chi-squared tests for dichotomous variables and Student's t-tests for continuous variables.

Unadjusted and adjusted regression analyses were performed; the adjusted analyses included race/ethnicity (non-Hispanic white, other), educational level (less than high school, high school diploma or some post high school, college degree or greater), and child's birth year. Race/ethnicity and educational level were included as routine potential confounders and because they differed between cases with and without controls. Other covariates, such as income level, paternal age at child's birth, marital status, and whether the father's interview was conducted by proxy, did not alter ORs by more than 10% and were therefore not included in the final models.

Simulations

In order to investigate the extent of matching of cases and controls, we compared the observed concordance of selected demographic and exposure variables between cases and controls with concordance rates that would be expected under random selection of controls without matching. We accomplished this by using simulated datasets generated by permutation methods. We simulated 1000 datasets in which the case- control status for each individual was held fixed but the binary variable of interest in the control group was randomly assigned in order to mimic random population ascertainment where the variable is unrelated to control selection. This approach kept the total number of controls with and without the exposure the same as in the original data, but it often changed the number of controls with the exposure within a case-control set.

Using these 1000 simulated datasets, we calculated 2 different measures of concordance. For the first measure, each case-control set was scored as concordant if all the controls had the same value of the binary demographic or exposure variable as the case. The proportion of all sets that were concordant by this definition was calculated and is referred to as the perfect concordance rate. In order to consider the variable number of controls per set, we calculated a second measure: the proportion of controls that had the same value of the variable as the case. The mean of this proportion was calculated across all sets and is referred to as the proportional concordance rate. Concordance rates calculated for the original study data were then compared to the simulated distributions of the concordance rates. Evidence of overmatching is suggested if the concordance rate computed from the original data lies in the extreme tails of the simulated concordance distribution.

Analyses were performed using SPSS (version 16.0, SPSS, Inc., Chicago, Illinois) and STATA/IC (version 10.0, Stata Corp., College Station, Texas) software. SAS version 9.1 (SAS Institute, Inc., Cary, North Carolina) was used for simulations and concordance calculations.

RESULTS

Control recruitment in the retinoblastoma study

We had planned to randomly select friends from the parents' lists. However, parents did not have contact information on all friends and were not willing to contact all friends for the study. Rather, parents were willing to contact for the study only friends that they chose and, therefore, we had to abandon random selection.

In this study, 199 interviewed case families were asked to nominate friends as potential controls. Of these, 163 (82%) nominated and provided contact information on at least 1 potential control. For 137 (69%) cases, the father of at least 1 control completed the father's interview directly or by proxy (Table 1); 56 cases had 1 control, 52 had 2 controls and 29 had 3 controls. Cases were slightly more likely to have an interviewed control mother than control father; 142 (71%) cases had an interviewed control mother.

Of the 62 case families without an interviewed control, 36 case families did not nominate a control or did not provide contact information for nominated controls. For the remaining 26 families without controls, the failure to find a control occurred after contact with the potential controls began; controls of 13 cases actively refused, 11 did not schedule or did not keep an interview appointment before the study ended, and 2 were ineligible due to the child's adoption or parents' not speaking English or Spanish.

The description above uses the case as the reference for determining response. However, for some cases, we attempted to contact additional controls who did not participate. Of the 374 potential friend controls we were able to contact, the fathers of 247 (66%) participated. The others were ineligible (n=12, 3%) due to the child's age or adoption, lack of phone service, or parents' language other than English or Spanish; refused (n=72, 19%); or could not be interviewed before the study ended (n=42, 11%). The response proportion for mothers was slightly higher at 70%.

Comparison of cases with and without controls

Cases with a control that completed the study interview and cases without an interviewed control differed markedly in demographic characteristics and smoking history (Table 2). Of the case fathers without controls, 37% were non-Hispanic white compared to 76% of cases with controls. Case fathers without controls also had less education, lower incomes, were less likely to be married, more likely to have smoked in the year before the pregnancy, and more likely to have been interviewed by proxy. Comparison of case and control fathers on other exposures, such as binge drinking, apple consumption, cured meat intake, body mass index, and father's age at the index child's birth, generally showed smaller differences.

Comparison of analytic approaches

Table 3 presents unadjusted and adjusted ORs for selected exposures determined by the 3 analytic approaches: matched, restricted unmatched, and full unmatched.

Overall, the unadjusted ORs from the matched and restricted unmatched analyses were similar, while those from the full unmatched analyses were substantially different. For example, for smoking, the ORs were 1.2 [95% CI 0.8, 2.0], 1.3 [95% CI 0.8, 2.2], and 2.0 [95% CI 1.3, 3.0] for the matched, restricted unmatched, and full unmatched analyses, respectively. Although the unadjusted ORs from matched and restricted unmatched analyses (i.e. the analyses that discarded cases without controls) were similar, there were slight to moderate differences. One of the largest differences occurred for father's age 30 with an OR of 1.7 [95% CI 0.9, 3.0] in the matched analysis compared to 1.3 [95% CI 0.8, 2.1] in the restricted unmatched analysis.

When adjusted for race/ethnicity, education, and birth year, the differences between the full unmatched ORs and those from the other 2 analyses were reduced compared to the unadjusted analyses; the differences were often minimal, but slight to moderate differences remained for some factors. For example, the ORs for smoking in the year before the pregnancy were 1.2 [95% CIs 0.7, 2.1 and 0.7, 2.0] for both the matched and restricted unmatched analyses and 1.5 [95% CI 1.0, 2.4] from the full unmatched analysis. In contrast, the ORs for multivitamin use differed minimally by analytic approach; the ORs were 1.5 [95% CI 0.9, 2.5], 1.5 [95% CI 0.9, 2.3], and 1.4 [95% CI 0.9, 2.1] for the matched, restricted unmatched, and full unmatched analyses, respectively.

We calculated ORs for race/ethnicity, educational level, and income to estimate their effect on disease. In both adjusted and unadjusted analyses, the ORs from the matched and restricted unmatched analyses were similar while those from the full unmatched analyses differed. Compared to the matched and restricted unmatched ORs, the full unmatched ORs

for race/ethnicity and education were more extreme and statistically significant. For income, a pattern was difficult to discern with the unadjusted full unmatched OR significantly less than 1.0 and the adjusted full unmatched OR close to 1.0. For age at child's birth, the results from adjusted matched, restricted unmatched, and full unmatched analyses did not differ appreciably.

Concordance rates

For the demographic factors studied, the perfect and proportional concordance rates in case-control sets in the original data exceeded the mean concordance rates from the simulated data in which random assignment of control exposure status was assumed (Figure 1A and 1B). The differences between the original and simulated concordances were generally large. For example, the perfect concordance rate for educational level was 60% in the original data and 32% in the simulated data. All of the observed rates fell outside or nearly outside the distribution of the simulated data.

For the exposures, the observed concordances exceeded the mean simulated concordances except for binge drinking (Figure 2A and 2B). The interquartile ranges of the simulated concordance distributions were narrow, for example from 0.35 to 0.40 for perfect concordance of multivitamin use. Thus, the observed concordances, which fell mostly beyond the 75th percentile and often beyond the 95th percentile of the simulated concordance distribution, did not differ greatly in magnitude from the mean simulated concordance. For example, the observed and mean simulated perfect concordances for smoking were 0.52 and 0.50, respectively. The largest difference was that for weekly apple consumption with observed and mean simulated proportional concordances of 0.58 and 0.49, respectively.

DISCUSSION

In our retinoblastoma study, we found the selection and recruitment of friend controls to be convenient and feasible. As presumed by many, the response rate was reasonably high, an appealing characteristic given that control recruitment is difficult. However, we were not able to follow the recommendation of randomly selecting controls from the case's friends because the cases chose who to name and who they were willing to contact and have us contact. Our combined group of friend and relative controls is comparable to many friend control groups described in the literature in that the control groups include non-biological relatives. In the literature, the most common non-biological relative used as a friend control is a spouse while in our study it was a brother-in-law.

Epidemiologists have assumed that the response rate among friend controls is higher than that among general population controls. Our experience and other reports generally confirm that the proportion of friends who participate is reasonably high, although without a direct comparison we cannot say our study response rate is higher than it would have been if we had attempted to obtain a population control group. Of potential controls for whom we had contact information, 66% completed interviews. Other researchers have reported response rates among friend controls of 63 – 100% for mailed questionnaires, mailed questionnaires with telephone follow-up, and in-person interviews,[7, 9, 16, 18, 19, 23] although the definition of the response rate is not always clear and may differ among reports. The response rate in this study, although reasonably high by today's standards, leaves considerable room for improvement. Although a detailed discussion is beyond the scope of this paper, researchers should consider attempts to increase the response rate, such as providing a small financial token prior to participation [25].

We did not anticipate the substantial proportion of cases, about 30%, for whom we were not able to recruit a control. In a search meant to be illustrative but not exhaustive, we identified 16 studies that used friend controls since 2000.[6–17, 19–21, 23] About one-third of these studies did not report the proportion of cases that nominated a control and/or had a control who completed participation. The proportion of cases with controls will vary depending on, for example, whether the control must be the same sex, the closeness of the age matching, and whether proxy interviews are acceptable. Thus, it is not appropriate to directly compare the proportion of cases with controls among studies. Nevertheless, it is worthwhile to note that the proportion of cases with a friend control who completed study participation ranged widely from 17 to 96%.[6, 8, 9, 11, 17, 19] Although we do not have specific recommendations, epidemiologists should consider efforts to increase the proportion of cases that nominate controls and, based on our results, target those efforts to cases of minority ethnicity and low SES. If successful, these efforts would improve the generalizability of results in addition to increasing the sample size of case-control sets.

In our study, cases without a matched control differed dramatically in race/ethnicity and SES from other cases. Few previous studies reported whether cases with and without friend controls differed. Worrall noted, as we did, that cases who did not nominate a control were much more likely to be non-white[22] and Kaplan et al. reported that cases without controls were older than cases with controls.[26] In addition to demographic characteristics, in our study, many but not all exposures differed substantially between cases with and without controls. In a study of melanoma in which all cases and controls were non-Hispanic white, Kanetsky et al. observed that cases with and without controls were generally similar in exposures and characteristics such as freckling and recreational sun exposure.[13] When we considered only non-Hispanic whites in our study, the differences between cases with and without controls were substantially smaller (data not shown) than in the whole study population. Thus, whether cases with and without friend controls differ in exposures may depend on the level of diversity in race/ethnicity and other characteristics in the study population.

Rothman and Greenland[4] recommend that when using a matched design, the matching factor(s) should be adjusted for in the analysis to avoid a biased estimate due to ascertainment; hence our matched analysis would be indicated. On the other hand, when using friend controls, the matching is not perfect since it is unintentional matching; hence a strong case could be made for conducting the unmatched restricted analyses as well. In our study, both analytic methods gave very similar results. However, from a practical standpoint, most epidemiologists would want to conduct the full unmatched analyses in order to be able to include all the participating cases, i.e. those with controls and those who do not have controls. When cases without controls were included and the analyses adjusted for the planned and inadvertent matching factors, the results for many exposures including binge drinking, multivitamin use, and paternal age were similar to those from the matched and restricted unmatched analyses. However, the results of the full unmatched analyses differed from the matched analyses for smoking. The direction of the differences was as expected from the demographic characteristics of the cases without controls. For example, as more of the cases without controls were of low SES which is associated with smoking, the OR for smoking in the full unmatched analysis was higher than the ORs from the other analyses. The exposures for which the ORs differed between the full unmatched and the other two analyses did not seem to follow a pattern. Multivitamin use and paternal age were associated with education but the ORs did not differ appreciably when cases without controls were included. Smoking status was also associated with education and the ORs differed depending on whether cases without controls were included. It seems that adjustment for the demographic factors did not completely eliminate the effect of the differences between cases with and without controls. Hence, we feel that the analyses that

include only cases with controls and their controls (matched or restricted unmatched) are the most appropriate. Excluding data on a substantial proportion of subjects, i.e. the cases without controls, is inefficient and the generalizability of the results may also suffer. In our study, the cases with and without controls differed demographically and excluding those without controls limits our ability to make inferences about individuals of minority ethnicity and lower SES. However, the data collected on the excluded cases may not have to be discarded as it may be used to address auxiliary aims, such as investigation of gene-environment interactions using case-case comparisons.

We compared observed concordance to that from simulated data that randomly permuted the controls' demographic factors and exposures. We interpreted the difference between the observed and simulated concordance rates as evidence for overmatching. However, without a randomly ascertained independent control group for comparison, we cannot definitively say that our friend controls are overmatched to the cases; instead, we can say they are highly concordant on several demographic characteristics, namely race/ethnicity, education, income, and paternal age. Perfectly concordant case-control sets do not contribute any information to the analysis of the concordant factor. Thus, the high concordance reduced our power to study the role of race/ethnicity, SES, and paternal age and one would not choose friend controls to assess the role of these factors on disease risk. However, one can argue that the inability to study the role of demographic characteristics is not a limitation of friend controls as case-control studies are not the ideal method for studying SES and race/ethnicity as risk factors because of the potential for response rates that differ by these factors. For example, population controls chosen by RDD and through birth certificates differed from the general population in race and SES, potentially biasing study results.[27] The high concordance of cases and friend controls on SES provides an advantage: any observed associations are unlikely to be confounded by SES.

Although cases and controls were highly concordant on demographic factors, we observed little evidence of greater than random case-control concordance for the exposures analyzed. This was true even for exposures such as smoking that were strongly associated with education and the other demographic factors. This observation suggests that high concordance for demographic factors does not necessarily result in increased concordance for exposures associated with those factors.

We cannot assess the validity of the observed results, i.e., whether the use of friend controls provided an unbiased estimate of the OR. The comparison of ORs observed with friends and those observed with another control group in the same study or with ORs observed in other studies with different designs or control groups might address the issue of validity. For retinoblastoma, few other data exist on possible risk factors with which to compare our results. However, the elevated OR we observed for paternal history of lower GI series is consistent with the result of a previous study that used RDD controls,[28] suggesting the validity of our results using friend controls for at least one exposure.

Friend controls are not ideal and may not work for every study. However, when population controls are not feasible or not appropriate for the study design, friend controls may represent a viable option. In our study of retinoblastoma, we observed friend controls to be a convenient source of SES-matched controls with an acceptable response rate and without notable overmatching on exposures. It seems likely that use of friend controls in other studies will require, as it did here, careful attention to the possibility of demographic differences between cases with and without controls and to the implications of such differences in choosing an analytic approach. Successful efforts to increase the proportion of cases nominating friend controls and the proportion of friend controls participating would improve the efficiency of these studies and the generalizability of the results.

Acknowledgments

References

1. Wacholder S, Silverman DT, McLaughlin JK, Mandel JS. Selection of controls in case-control studies. II. Types of controls. American Journal of Epidemiology. 1992; 135:1029–1041. [PubMed: 1595689]

2. Rothman, KJ.; Greenland, S. Modern Epidemiology. Phlladelphia: Lippincott Williams & Wilkins; 1998. Case-Control Studies; p. 93-114.

3. Zondervan KT, Cardon LR, Kennedy SH. What makes a good case-control study? Design issues for complex traits such as endometriosis. Human Reproduction. 2002; 17:1415–1423. [PubMed: 12042253]

4. Rothman, KJ.; Greenland, S. Modern Epidemiology. Philadelphia: Lippincott Williams & Wilkins; 1998. Matching; p. 147-162.

5. Bunin GR, Spector LG, Olshan AF, Robison LL, Roesler M, Grufferman S, et al. Secular Trends in Response Rates for Controls Selected by Random Digit Dialing in Childhood Cancer Studies: A Report from the Children's Oncology Group. American Journal of Epidemiology. 2007; 166:109–116. [PubMed: 17456476]

6. Amend KL, Elder JT, Tomsho LP, Bonner JD, Johnson TM, Schwartz J, et al. EGF gene polymorphism and the risk of incident primary melanoma. Cancer Research. 2004; 64:2668–2672. [PubMed: 15087376]

7. Chaudhari M, Moroldo MB, Shear E, Hillard P, Thompson SD, Lan D, et al. Impaired reproductive fitness in mothers of children with juvenile autoimmune arthropathies. Rheumatology (Oxford). 2006; 45:1282–1287. [PubMed: 16595515]

8. Cust AE, Schmid H, Maskiell JA, Jetann J, Ferguson M, Holland EA, et al. Population-based, case-control-family design to investigate genetic and environmental influences on melanoma risk: Australian Melanoma Family Study. American Journal of Epidemiology. 2009; 170:1541–1554. [PubMed: 19887461]

9. Hilliquin P, Allanore Y, Coste J, Renoux M, Kahan A, Menkes CJ. Reduced incidence and prevalence of atopy in rheumatoid arthritis. Results of a case-control study. Rheumatology (Oxford). 2000; 39:1020–1026. [PubMed: 10986309]

10. Hodgson DC, Pintilie M, Gitterman L, Dewitt B, Buckley CA, Ahmed S, et al. Fertility among female hodgkin lymphoma survivors attempting pregnancy following ABVD chemotherapy. Hematology Oncology. 2007; 25:11–15.

11. Il'yasova D, McCarthy B, Marcello J, Schildkraut JM, Moorman PG, Krishnamachari B, et al. Association between glioma and history of allergies, asthma, and eczema: a case-control study with three groups of controls. Cancer Epidemiology Biomarkers & Prevention. 2009; 18:1232–1238.

12. Kanetsky PA, Holmes R, Walker A, Najarian D, Swoyer J, Guerry D, et al. Interaction of glutathione S-transferase M1 and T1 genotypes and malignant melanoma. Cancer Epidemiology Biomarkers & Prevention. 2001; 10:509–513.

13. Kanetsky PA, Panossian S, Elder DE, Guerry D, Ming ME, Schuchter L, et al. Does MC1R genotype convey information about melanoma risk beyond risk phenotypes? Cancer. 116:2416–2428. [PubMed: 20301115]

14. Kirsch R, Wirrell E. Do cognitively normal children with epilepsy have a higher rate of injury than their nonepileptic peers? Journal of Child Neurology. 2001; 16:100–104. [PubMed: 11292213]

15. Li H, Wetten S, Li L, St Jean PL, Upmanyu R, Surh L, et al. Candidate single-nucleotide polymorphisms from a genomewide association study of Alzheimer disease. Archives of Neurology. 2008; 65:45–53. [PubMed: 17998437]

16. Parikh-Patel A, Gold E, Utts J, Gershwin ME. The association between gravidity and primary biliary cirrhosis. Annals of Epidemiology. 2002; 12:264–272. [PubMed: 11988415]

17. Peeters AC, van Landeghem BA, Graafsma SJ, Kranendonk SE, Hermus AR, Blom HJ, et al. Low vitamin B6, and not plasma homocysteine concentration, as risk factor for abdominal aortic aneurysm: a retrospective case-control study. Journal of Vascular Surgery. 2007; 45:701–705. [PubMed: 17398378]

18. Sigurdson AJ, Chang S, Annegers JF, Duphorne CM, Pillow PC, Amato RJ, et al. A case- control study of diet and testicular carcinoma. Nutrition and Cancer. 1999; 34:20–26. [PubMed: 10453437]

19. Svalheim S, Tauboll E, Bjornenak T, Roste LS, Morland T, Saetre ER, et al. Do women with epilepsy have increased frequency of menstrual disturbances? Seizure. 2003; 12:529–533. [PubMed: 14630488]

20. Walcott FL, Hauptmann M, Duphorne CM, Pillow PC, Strom SS, Sigurdson AJ. A case-control study of dietary phytoestrogens and testicular cancer risk. Nutrition and Cancer. 2002; 44:44–51. [PubMed: 12672640]

21. Wells PS, Anderson JL, Rodger MA, Carson N, Grimwood RL, Doucette SP. The factor XIII Val34Leu polymorphism: is it protective against idiopathic venous thromboembolism? Blood Coagulation & Fibrinolysis. 2006; 17:533–538. [PubMed: 16988547]

22. Worrall BB, Brown DL, Brott TG, Brown RD, Silliman SL, Meschia JF. Spouses and unrelated friends of probands as controls for stroke genetics studies. Neuroepidemiology. 2003; 22:239–244. [PubMed: 12792144]

23. Zhou W, Liu G, Thurston SW, Xu LL, Miller DP, Wain JC, et al. Genetic polymorphisms in N-acetyltransferase-2 and microsomal epoxide hydrolase, cumulative cigarette smoking, and lung cancer. Cancer Epidemiology Biomarkers & Prevention. 2002; 11:15–21.

24. Dryja TP, Morrow JF, Rapaport JM. Quantification of the paternal allele bias for new germline mutations in the retinoblastoma gene. Human Genetics. 1997; 100:446–449. [PubMed: 9272170]

25. Dillman, D. Mail and internet surveys: the tailored design method. 2. Hoboken: Wiley; 2000.

26. Kaplan S, Novikov I, Modan B. A methodological note on the selection of friends as controls. International Journal of Epidemiology. 1998; 27:727–729. [PubMed: 9758132]

27. Puumala SE, Spector LG, Robison LL, Bunin GR, Olshan AF, Linabery AM, et al. Comparability and representativeness of control groups in a case-control study of infant leukemia: a report from the Children's Oncology Group. American Journal of Epidemiology. 2009; 170:379–387. [PubMed: 19498073]

28. Bunin GR, Meadows AT, Emanuel BS, Buckley JD, Woods WG, Hammond GD. Pre- and post-conception factors associated with heritable and non-heritable retinoblastoma. Cancer Research. 1989; 49:5730–5735. [PubMed: 2790788]
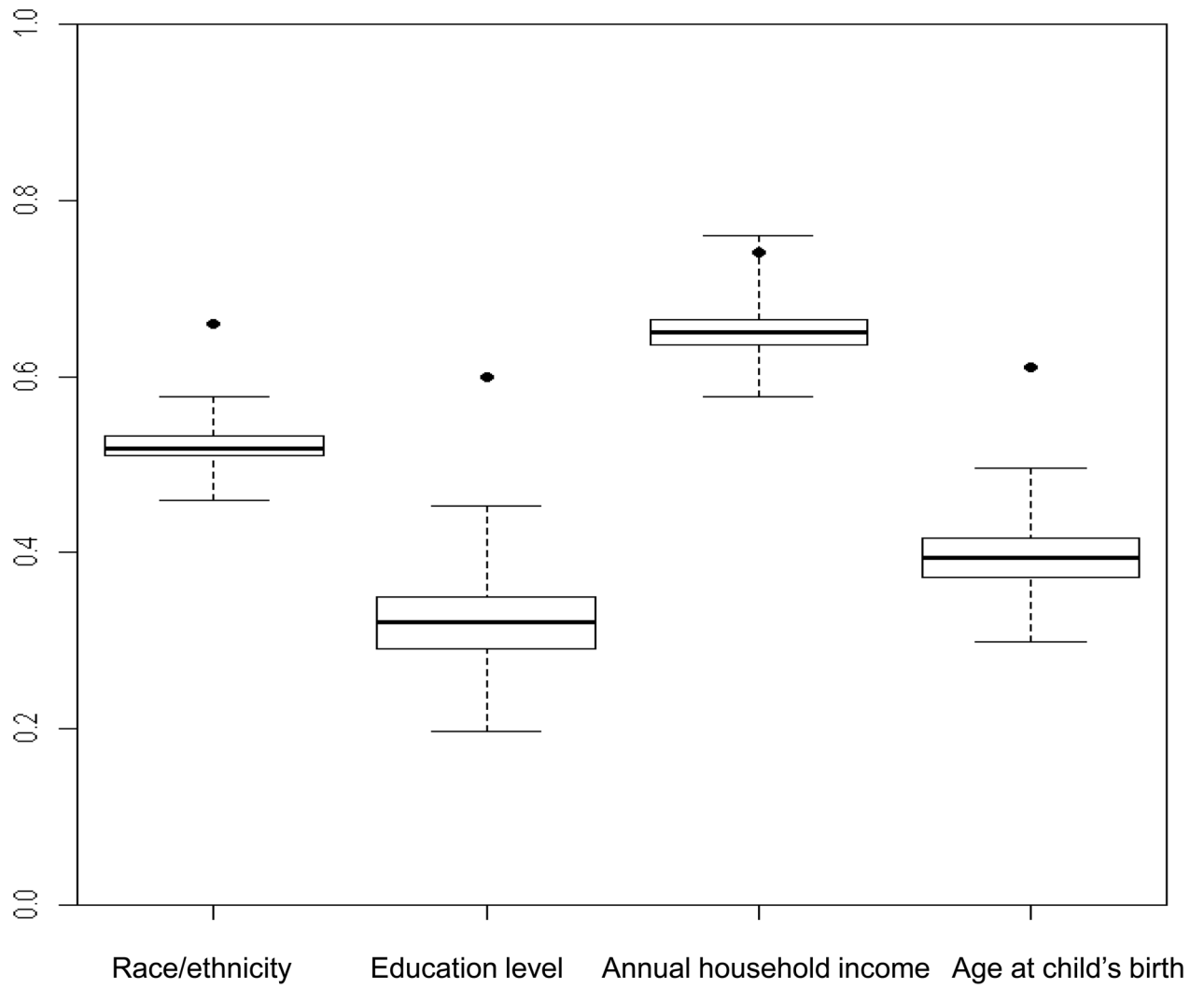
Figure 1. Box-and-whisker plots show distribution of concordance rates from simulated data for cases and friend controls within case-control sets (n=137) for selected demographic characteristics

Boxes show the median and interquartile range of the concordance rates from simulated data; the vertical extensions indicate the minimum and maximum values. Dots show the observed concordance rate from the case-control study. The demographic characteristics were dichotomized as follows: race/ethnicity (non-Hispanic white, other); education level (<college graduate, college graduate); annual household income (<$35,000, $35,000); age at child's birth (<30 years, 30 years).

A. Perfect concordance rates

From left to right, the concordances from the original data (percentile values) are: 0.66 (100), 0.60 (100), 0.74 (99.8) and 0.61(100). From left to right, the mean concordances from the simulated data are: 0.52, 0.32, 0.65 and 0.40.

B. Proportional concordance rates

From left to right, the concordances from the original data (percentile values) are: 0.86 (100), 0.70 (100), 0.67 (100) and 0.72 (100). From left to right, the mean concordances from the simulated data are: 0.66, 0.49, 0.56 and 0.55.
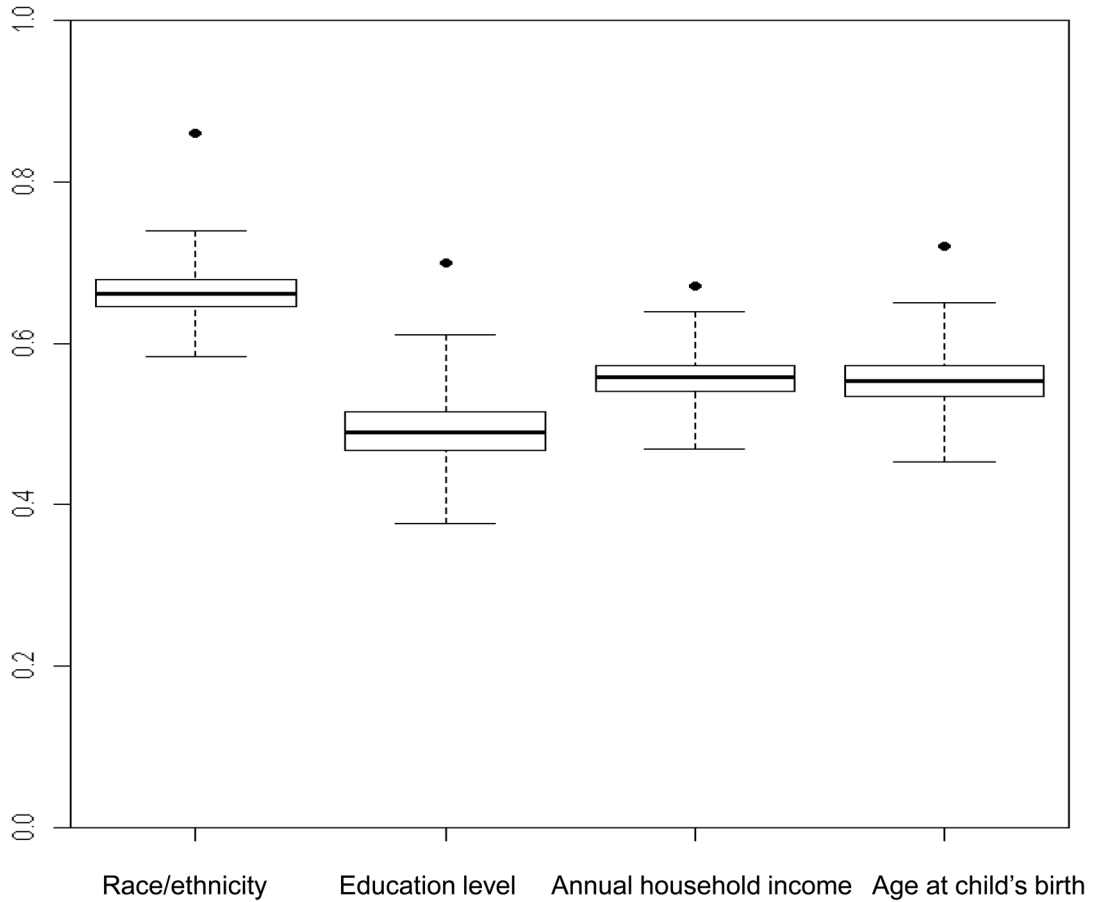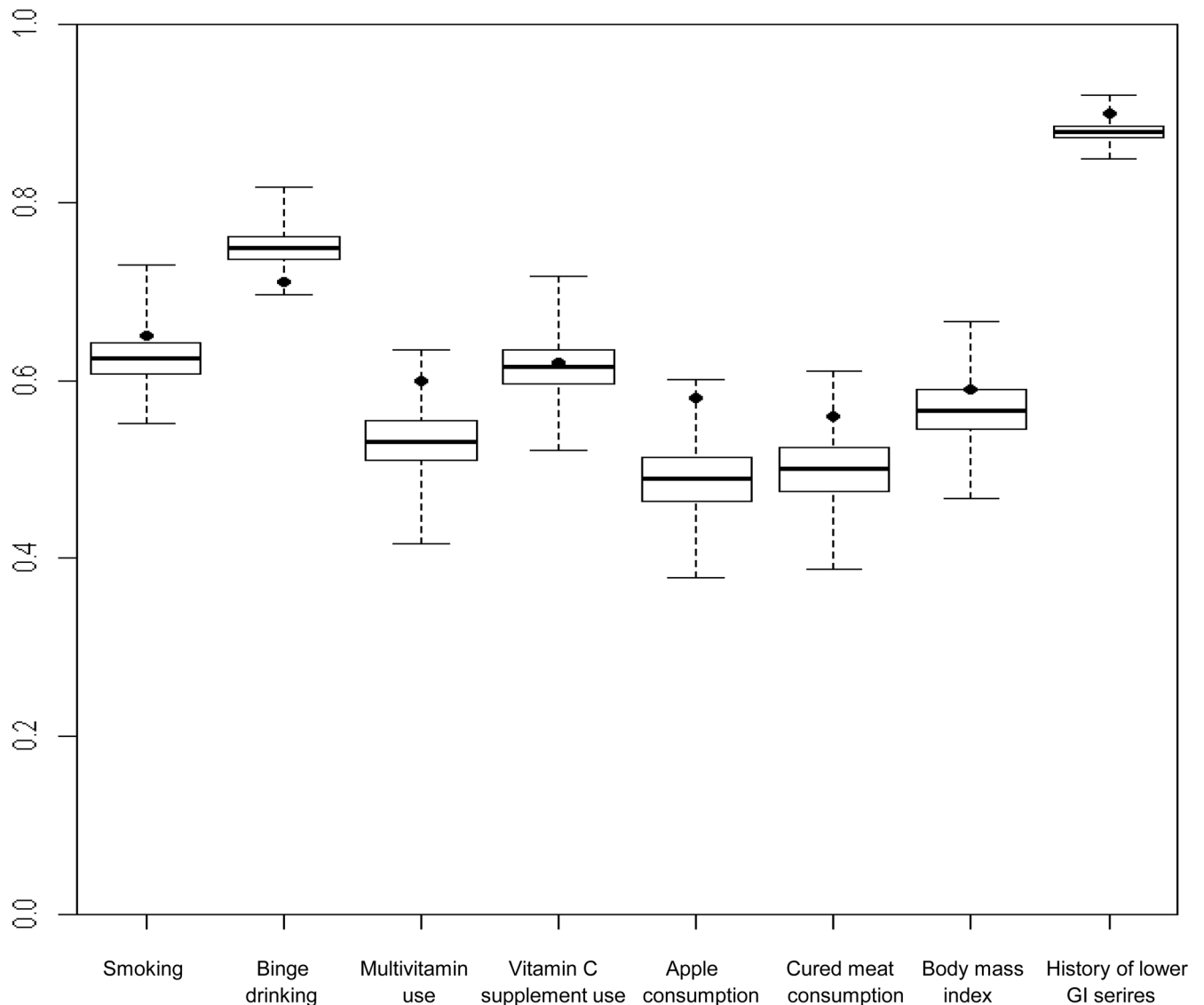
Figure 2. Box-and-whisker plots show distribution of concordance rates from simulated data for cases and friend controls within case-control sets (n=137) for selected exposures
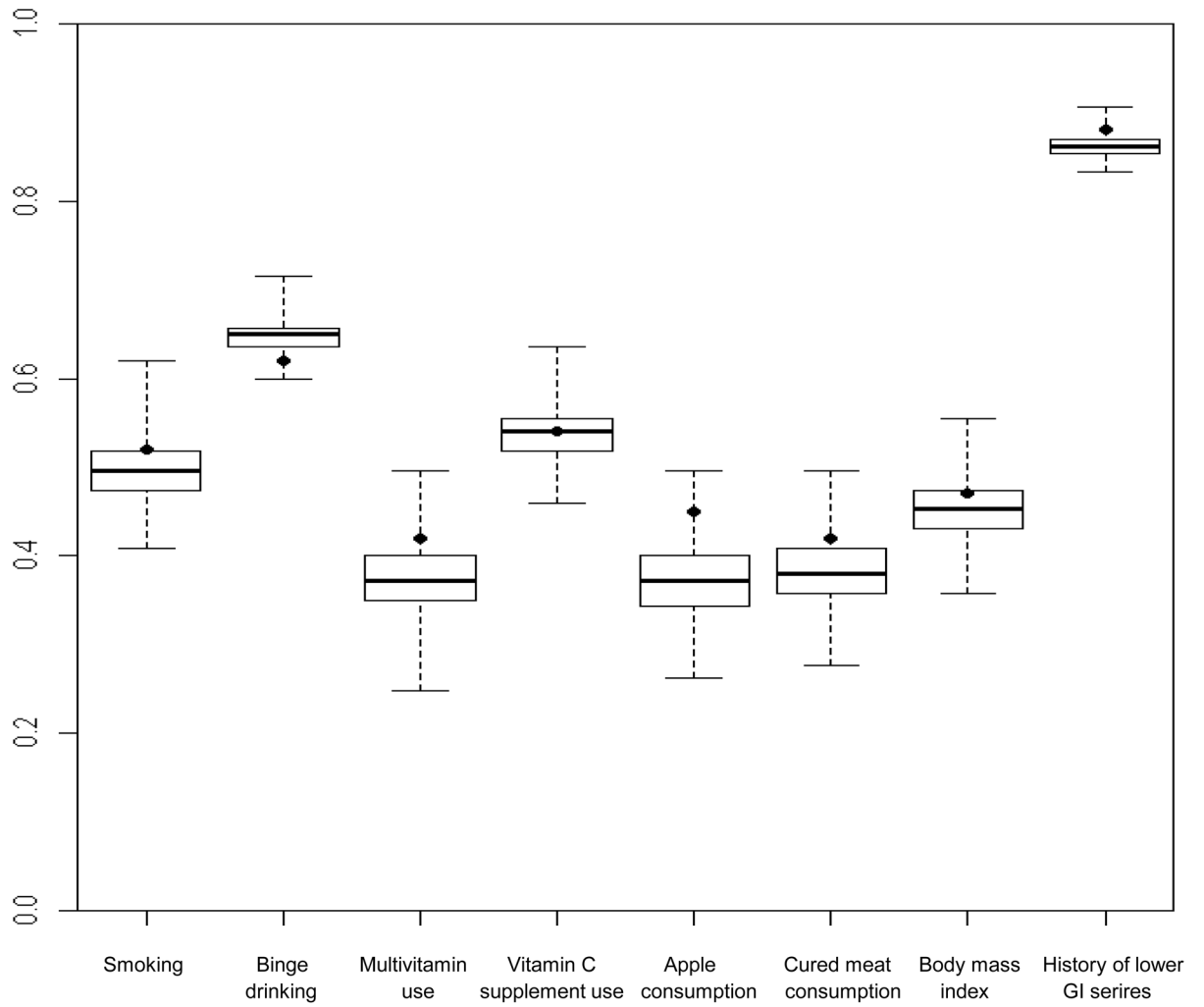Boxes show the median and interquartile range of the concordance rates from simulated data; the vertical extensions indicate the minimum and maximum values. Dots show the observed concordance rate from the case-control study. The exposures were dichotomized as follows: smoking (yes, no); binge drinking (<1 time/month, 1 time/month); multivitamin use (yes, no); vitamin C supplement use (yes, no); apple consumption (<1/week, 1/week); cured meat consumption (2 servings/week, >2 servings/week); body mass index (<25, 25); history of lower gastrointestinal series (yes, no).

A. Perfect concordance rates
From left to right, the concordances from the original data (percentile values) are: 0.52 (81), 0.62 (6), 0.42 (89), 0.54 (48), 0.45 (97), 0.41(80), 0.47(69) and 0.88 (95). From left to right, the mean concordances from the simulated data are: 0.50, 0.62, 0.37, 0.54, 0.37, 0.38, 0.45 and 0.86.

B. Proportional concordance rates

From left to right, the concordances from the original data (percentile values) are: 0.65 (84), 0.71 (1), 0.60 (98), 0.62 (57), 0.58 (99), 0.56 (95), 0.59 (75) and 0.90 (97). From left to right, the mean concordances from the simulated data are: 0.62, 0.75, 0.53, 0.62, 0.49, 0.50, 0.57 and 0.88.

Table 1

Outcome of Selection of Friend Control Fathers for 199 Case Fathers in a Case-control Study of Retinoblastoma, United States, 2002 – 2007

	Any control	
	#	%
Total cases	**199**	**100**
At least 1 control interviewed	**137**	**69**
Case did not nominate control and provide correct contact information	**36**	**18**
Case declined to nominate control	7	4
Case had no eligible friend	17	9
Case nominated control but never provided contact information	11	6
Case provided Incorrect contact information	1	1
Contacted control did not participate	**26**	**13**
Control ineligible	2	1
Control refused	13	7
Father's interview not completed before study ended	11	6

Table 2

Fathers' Demographic, Exposure, and Interview Characteristics: Cases Without Controls, Cases With Controls, and Controls From a Case-control Study of Retinoblastoma Using Friend Controls, United States, 2002 – 2007.

Exposure/characteristic[a]	Cases without controls N=62[b]		Cases with controls N=137[b]		Controls N=247[b]	
	#	%	#	%	#	%
Race/ethnicity[c]						
White, non-Hispanic	23	37	104	76	202	82
Black, non-Hispanic	13	21	17	12	19	8
Hispanic	15	24	10	7	18	7
Other	11	18	6	4	8	3
Education						
Less than high school	14	23	5	4	13	5
High school graduate	18	29	36	26	45	18
Some post high school	17	27	33	24	51	21
College graduate	8	13	39	29	88	36
Graduate/professional school	5	8	23	17	50	20
Income						
< $25,000	18	37	8	7	21	10
$25 – 35,000	8	17	13	11	19	9
$35 – 50,000	5	10	26	21	37	17
$50 – 75,000	11	23	28	23	50	23
> $75,000	6	12	47	39	86	40
Age at child's birth						
< 25	8	13	11	8	18	7
25 – 29	19	31	30	22	71	29
30 – 34	13	21	43	31	82	33
35 – 39	16	26	40	29	53	21
40 +	5	8	13	9	23	9
Marital status at interview						
Married	44	72	125	91	229	93
Not married	17	28	12	9	16	7

Exposure/characteristic[a]	Cases without controls N=62[b]		Cases with controls N=137[b]		Controls N=247[b]	
	#	%	#	%	#	%
Interviewed by proxy						
Yes	17	27	16	12	43	17
No	45	73	121	88	204	83
Smoked						
Yes	34	55	38	28	55	22
No	28	45	99	72	192	78
# cigarettes/day						
None	28	47	99	72	192	78
1–10/day	18	30	19	14	31	13
11–50/day	14	23	19	14	24	10
Binge drinking[d]						
<1time/month	47	80	119	87	207	84
1time/month	12	20	18	13	40	16
Multivitamin use						
Yes	17	28	55	40	83	34
No	44	72	82	60	164	66
Vitamin C supplement use						
Yes	8	15	38	31	49	22
No	44	85	84	69	174	78
Apple consumption						
<1 serving/week	33	63	75	58	100	43
1 serving/week	19	36	54	42	130	57
Cured meat intake						
<2 servings/week	28	54	47	39	116	52
2 servings/week	24	46	75	61	107	48
Body mass index						
<25	18	35	35	29	76	34
25	33	65	87	71	147	66
History of lower GI[e] series						
Yes	3	6	11	9	8	4

Exposure/characteristic[a]	Cases without controls N=62[b]		Cases with controls N=137[b]		Controls N=247[b]	
	#	%	#	%	#	%
No	49	94	111	91	215	96

[a] Exposures/characteristics pertain to the father in the year before the index pregnancy except that the time period for history of lower GI series is any time prior to the index child's conception.

[b] Missing data explain the differing number of total subjects among exposures/characteristics. A small proportion of fathers completed shorter questionnaires that did included a shortened supplement section and no diet or medical radiation section. Thus, for vitamin C supplement use, cured meat intake, apple consumption, body mass index, and history of lower GI series, the maximum numbers are 52 cases without controls, 122 cases with controls, and 223 controls.

[c] Cases with and without controls differ significantly in race/ethnicity ($P<0.001$), educational level ($P<0.001$), income ($P<0.001$), marital status ($P=0.001$), father interviewed by proxy ($P=0.008$), smoked ($P<0.001$), and # cigarettes/day ($P=0.002$).

[d] Binge drinking was defined as consuming 6 or more drinks on one occasion

[e] Gastrointestinal

Table 3

Unadjusted and Adjusted Odds Ratios Using 3 Analytic Approaches: Matched (Matched Analysis of Case-control Sets), Restricted Unmatched (Unmatched Analysis of Members of Case-control Sets), and Full Unmatched (Unmatched Analysis of All Cases and Controls); Case-control Study of Retinoblastoma Using Friend Controls, United States, 2002 – 2007.

Exposure/characteristic[d]	Unadjusted odds ratios and 95% confidence intervals						Adjusted[a] odds ratios and 95% confidence intervals					
	Matched		Restricted unmatched		Full unmatched		Matched		Restricted unmatched		Full unmatched	
	OR[b]	95% CI[c]	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI
Smoked in year before pregnancy												
Ever	1.2	0.8, 2.0	1.3	0.8, 2.2	2.0**	1.3, 3.0	1.2	0.7, 2.1	1.2	0.7, 2.0	1.5	1.0, 2.4
# cigarettes												
None (Reference)												
1–10	1.1	0.6, 2.1	1.2	0.6, 2.2	1.8*	1.1, 3.1	1.0	0.5, 2.1	1.0	0.5, 2.2	1.2	0.6, 2.3
11+	1.4	0.7, 2.8	1.5	0.8, 2.9	2.1*	1.2, 3.7	1.5	0.8, 2.9	1.1	0.4, 2.6	1.7	0.8, 3.6
Binge drinking ≥1 time/month[e]	0.7	0.4, 1.3	0.8	0.4, 1.4	0.9	0.5, 1.5	0.7	0.4, 1.3	0.7	0.4, 1.4	0.8	0.5, 1.3
Multivitamin use	1.5	0.9, 2.4	1.3	0.9, 2.0	1.1	0.8, 1.7	1.5	0.9, 2.5	1.5	0.9, 2.3	1.4	0.9, 2.1
Vitamin C supplement use	1.8*	1.1, 3.1	1.6	1.0, 2.6	1.2	0.8, 1.9	2.0*	1.2, 3.5	1.8*	1.1, 3.0	1.5	0.9, 2.4
≥1 apple serving/week	0.5*	0.3, 0.9	0.6**	0.4, 0.9	0.5**	0.4, 0.8	0.5*	0.3, 0.9	0.6*	0.4, 0.9	0.6**	0.4, 0.9
≥2 cured meat servings/week	1.8*	1.1, 2.8	1.7*	1.1, 2.7	1.4	0.9, 2.1	1.8*	1.1, 2.9	1.7*	1.1, 2.7	1.4	0.9, 2.1
Body mass index ≥25	1.2	0.7, 2.0	1.3	0.8, 2.1	1.2	0.8, 1.8	1.2	0.7, 2.1	1.3	0.8, 2.1	1.2	0.8, 1.8
History of lower GI series	3.5*	1.2, 10.7	2.7*	1.0, 6.8	2.4*	1.0, 5.9	3.9*	1.3, 12.0	3.0*	1.1, 7.8	2.9*	1.2, 7.2
Race/ethnicity other than non-Hispanic white	1.0	0.4, 2.3	1.4	0.9, 2.4	2.6***	1.7, 4.0	1.0	0.4, 2.3	1.2	0.7, 2.1	2.2**	1.4, 3.4
College graduate or higher	0.8	0.4, 1.3	0.7	0.4, 1.0	0.5***	0.3, 0.7	0.8	0.4, 1.3	0.7	0.4, 1.0	0.6**	0.4, 0.8
Income ≥$35,000	1.4	0.7, 3.0	1.1	0.6, 2.0	0.6*	0.4, 0.9	1.5	0.7, 3.3	1.7	0.8, 3.3	1.1	0.6, 2.0

| | Unadjusted odds ratios and 95% confidence intervals | | | | | | Adjusted[a] odds ratios and 95% confidence intervals | | | | | |
| | Matched | | Restricted unmatched | | Full unmatched | | Matched | | Restricted unmatched | | Full unmatched | |
	OR[b]	95% CI[c]	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI
Age 30 at child's birth	1.7	0.9, 3.0	1.3	0.8, 2.1	1.1	0.7, 1.6	1.6	0.9, 2.9	1.6	1.0, 2.5	1.5	1.0, 2.3

*
p<0.05

**
p<0.01

***
p<0.001

[a] Odds ratios are adjusted for child's birth year, father's race/ethnicity (non-Hispanic white, other) and father's educational level (< high school, high school/some post high school, college graduate) except as follows; the ORs for demographic characteristics are adjusted for the other demographic characteristics, i.e., the OR for non-white race/ethnicity is adjusted for birth year and father's educational level, the OR for college graduate or higher is adjusted for birth year and father's race/ethnicity, and the OR for income $35,000 is adjusted for father's race/ethnicity and father's educational level

[b] OR odds ratio

[c] CI confidence interval

[d] Exposures/characteristics pertain to the father in the year before the index pregnancy except that the time period for history of lower GI series is any time prior to the index child's conception.

[e] Binge drinking was defined as consuming 6 or more drinks on one occasion