NIH Public Access
Author Manuscript
Cancer Epidemiol Biomarkers Prev. Author manuscript; available in PMC 2013 June 01.

Published in final edited form as:
Cancer Epidemiol Biomarkers Prev. 2012 June ; 21(6): 881–886. doi:10.1158/1055-9965.EPI-11-1066.

Selection Bias in Population-Based Cancer Case–Control Studies Due to Incomplete Sampling Frame Coverage

Matthew C. Walsh[1,2], **Amy Trentham-Dietz**[1,2], **Ronald E. Gangnon**[2], **F. Javier Nieto**[2], **Polly A. Newcomb**[1,3], and **Mari Palta**[2]

[1]Paul P. Carbone Comprehensive Cancer Center, University of Wisconsin-Madison, Madison, Wisconsin

[2]Department of Population Health Sciences, University of Wisconsin-Madison, Madison, Wisconsin

[3]Cancer Prevention Program, Fred Hutchinson Cancer Research Center, Seattle, Washington

Abstract

Background—Increasing numbers of individuals are choosing to opt out of population-based sampling frames due to privacy concerns. This is especially a problem in the selection of controls for case–control studies, as the cases often arise from relatively complete population-based registries, whereas control selection requires a sampling frame. If opt out is also related to risk factors, bias can arise.

Methods—We linked breast cancer cases who reported having a valid driver's license from the 2004–2008 Wisconsin women's health study (N = 2,988) with a master list of licensed drivers from the Wisconsin Department of Transportation (WDOT). This master list excludes Wisconsin drivers that requested their information not be sold by the state. Multivariate-adjusted selection probability ratios (SPR) were calculated to estimate potential bias when using this driver's license sampling frame to select controls.

Results—A total of 962 cases (32%) had opted out of the WDOT sampling frame. Cases age <40 (SPR = 0.90), income either unreported (SPR = 0.89) or greater than $50,000 (SPR = 0.94), lower parity (SPR = 0.96 per one-child decrease), and hormone use (SPR = 0.93) were significantly less likely to be covered by the WDOT sampling frame (α = 0.05 level).

© 2012 American Association for Cancer Research.

Corresponding Author: Matthew C. Walsh, University of Wisconsin-Madison, Room 18, 6602 University Ave., Middleton, WI 53562. Phone: 608-821-1268; Fax: 608-821-1244; walsh2@wisc.edu.

Disclosure of Potential Conflicts of Interest
No potential conflicts of interest were disclosed.

Authors' Contributions
Conception and design: M.C. Walsh, A. Trentham-Dietz, F.J. Nieto
Development of methodology: M.C. Walsh, A. Trentham-Dietz, M. Palta
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): M.C. Walsh, A. Trentham-Dietz, P.A. Newcomb
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): M.C. Walsh, A. Trentham-Dietz, R.E. Gangnon, M. Palta
Writing, review, and/or revision of the manuscript: M.C. Walsh, A. Trentham-Dietz, R.E. Gangnon, F.J. Nieto, P.A. Newcomb, M. Palta
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): M.C. Walsh, A. Trentham-Dietz
Study supervision: A. Trentham-Dietz, M. Palta

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Conclusions—Our results indicate the potential for selection bias due to differential opt out between various demographic and behavioral subgroups of controls. As selection bias may differ by exposure and study base, the assessment of potential bias needs to be ongoing.

Impact—SPRs can be used to predict the direction of bias when cases and controls stem from different sampling frames in population-based case–control studies.

Introduction

Selection bias in population-based cancer research can affect the validity of the evidence used for public health practice (1). In epidemiologic case–control studies, cases and controls should arise from the same study base (2–5). However, the nature of case–control studies may lead to different sampling frames being used for cases and controls. For example, cases may arise from physician reports in a geographic region, whereas controls are selected from some type of list of the population in the same region. This is the situation when a statewide cancer registry is used to sample cases and a state-provided master file of licensed drivers is used to sample controls. Such studies make the assumption that coverage of the study base is comparable for both the case and control sampling frames. If the study base is all residents in a geographical area, then the case sampling frame needs to include all incident cases in that area, and the control sampling frame needs to include all residents in the same area. Bias will arise if sampling frame coverage of the study base is associated differently with risk factors among cases and controls. Inherently, there is no association of coverage with risk factors among cases, when a registry is nearly complete. If, however, no complete sampling frame exists for controls, bias will then arise if coverage of the control sampling frame is associated with risk factors. The potential for bias varies by exposure, sampling frame, and geographical setting requiring a fresh evaluation for most studies. While response rates and nonresponse error have been discussed extensively in the epidemiologic literature (6–10), the effects of the completeness of sampling frames have received little attention.

An official master file of drivers with valid licenses is available for epidemiologic studies in many states (11). Although coverage may vary, such files have historically been sufficiently complete to constitute a useful sampling frame (12–13). The largest coverage problems have resulted from (i) individuals who are not old enough to drive, (ii) those who have stopped renewing, or (iii) those who never obtained a driver’s license. More recently, legal constraints have compromised the coverage of these files. To comply with the Driver’s Privacy Protection Act (DPPA: 18 U.S.C. 2721–2725), the state of Wisconsin has instituted an "opt-out" program allowing drivers the option of keeping driver’s license data confidential and unavailable to research organizations. Using data from a population-based breast cancer case–control study and the Wisconsin Department of Transportation (WDOT),we had an opportunity to examine the effects of "opt out" due to privacy concerns affecting the sampling frame used to select population-based controls on the validity of study results.

Materials and Methods

Wisconsin women’s health study

This analysis used data from the breast cancer cases enrolled in the Wisconsin women’s health study (WWHS), a federally funded population-based case–control study designed to examine the associations of lifestyle factors and genetics with breast cancer risk (14). WWHS cases were incident breast cancer cases reported to the mandatory statewide cancer registry and interviewed for our study between 2004 and 2008. Data from the statewide cancer registry met high completeness standards (90%–95% complete). Some issues with

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

completeness due to failure of neighboring states in reporting cases have been previously addressed (15). Cases (N = 2,988; 74% response rate) completed a 35-minute structured telephonic interview. In addition to detailed questions concerning medication use, cases were queried about demographic characteristics, reproductive history, personal and family cancer history, physical activity, smoking, and alcohol consumption. Controls in the parent study were identified using a master file compiled by the WDOT of all licensed drivers, excluding the individuals who "opted out." This investigation used only data from the WWHS cases to investigate selection bias due to incomplete control sampling frame coverage.

Linkage

We matched the participating 2,988 WWHS breast cancer cases to WDOT driver's license files of individuals who had not "opted out" to estimate the completeness of the driver's license sampling frame used to select controls. WWHS cases were linked with data from the 2006 WDOT master file. A dichotomous variable was created indicating whether each WWHS case record matched a record from the driver's license master file. First, exact match linkages were conducted on the basis of gender, last name, first name, date of birth, and zip code between WWHS case records and WDOT driver's license records. Second, manual review of exact match linkages was conducted based first on gender, last name, first name, and date of birth and then based on gender, last name, and date of birth for all remaining unmatched case records. Manual review of the remaining unmatched WWHS cases, aided by various weighting schemes developed by the National Center for Health Statistics for the National Death Index, was then used to find any additional cases with matching WDOT license records (16).

Selection probability ratio calculation

It has been shown that the ratio of the probability of being on the sampling frame among individuals with a risk factor divided by the corresponding probability among those without the risk factor is a measure of the bias in the estimated odds ratio (OR) of disease for the factor (17–19). We will use such ratios to estimate the magnitude of bias that may be expected. For a given risk factor under investigation in a case–control study, let α denote the probability of a case with the risk factor being on the sampling frame used for cases, and let β denote the probability of a case without the risk factor being on the sampling frame used for cases. Let γ and δ denote the corresponding probabilities among controls of being on the sampling frame used to select controls, and OR, the true crude OR between the risk factor and disease. Previous work on bias focused on the effects of participation bias, but applies to the situation when lack of participation is caused by absence from the sampling frame used to select either cases or controls. This work has shown that the estimated OR under the scenario described is $[(\alpha/\beta)/(\gamma/\delta)] \times OR$; thus $[(\alpha/\beta)/(\gamma/\delta)]$ is a measure of bias. It was also shown that the same principle can be applied to an adjusted OR, using adjusted estimates of (α/β) and (γ/δ) obtained by averaging corresponding stratum specific quantities with the weights used in averaging the stratum-specific log ORs (19–20). It should be noted that if $(\alpha/\beta) = (\gamma/\delta)$, that is, risk factors affect coverage the same way among cases and controls, there is no bias. However, in our situation using data from the WWHS, case reporting to the Wisconsin Cancer Reports System is nearly 100% complete, so $\alpha/\beta = 1$. This implies there will be bias if $\gamma/\delta \neq 1$, that is, if risk factors are associated with control sampling frame coverage.

As risk factor information is not available for controls who are not on the drivers' license data frame, it is not possible to directly estimate γ/δ for the controls. Under the assumption that opting out patterns are similar for cases and controls, however, γ/δ is the ratio associated with opt out among both cases and controls, and we can estimate γ/δ from our

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

match of cases to the drivers' license file. The selection probability ratio (SPR) $1/(\gamma/\delta)$ can then be used to estimate the magnitude of bias. It should be noted that this is using data on a subset of cases solely for the estimation of bias in the control group, whereas all cases will be included in the final case–control analysis. Figure 1 illustrates how various SPR values might affect a study's results.

Statistical analysis

Analyses were conducted with SAS version 9.1 (SAS Institute). Among the 2,988 WWHS cases, we modeled probabilities of cases being on the sampling frame provided in 2006 by the WDOT on the basis of demographics (e.g., age, education, and income), lifestyle characteristics (e.g., physical activity, obesity, and alcohol intake), and medical factors (e.g., cancer screening and comorbidities). Multivariable-adjusted SPRs were estimated by fitting a generalized linear model with the log link function, poisson distribution, and robust error variance to the WWHS cases (17). Demographic factors (race, income, education, age, and marital status) and breast cancer risk factors (age at first birth, parity, antidepressant medication use, hormone replacement therapy, oral contraceptive use, body mass index at time of diagnosis, smoking status, total alcohol consumption, menopause status, family history of cancer, and age at menarche) were included in the initial models. While the demographic factors listed earlier were retained in all models, other breast cancer risk factors with P values greater than 0.20 were removed sequentially to obtain the final sampling frame coverage model.

The approach here to obtain ORs for specific risk factors and sampling ratios for coverage probabilities from separate regression analyses does not fulfill the criterion of the same weighting across adjustment variables or strata. Hence, the sampling ratios cannot be used to directly correct the ORs (18). Nonetheless, the adjusted SPRs provide guidance as to the expected magnitude of bias, and will be similar to those desired for correction if interaction effects are insignificant in the disease model, thereby minimizing the influence of the choice of averaging weights.

Results

The WDOT master file of licensed drivers for November 2006 included 3,018,192 records. This master file was linked with the 2,988 breast cancer cases that participated in the WWHS. A total of 2,026 (67.8%) of the WWHS cases were found to have a matching record on the WDOT master file. Of these 2,026 cases, 1,477 (72.9%) had an exact match based on gender, last name, first name, date of birth, and zip code. An additional 391 cases (19.3%) were matched on the basis of gender, last name, first name, and date of birth, and 66 (3.3%) were matched on gender, last name, and date of birth. Manual review, aided by weighting schemes based on those used by the National Death Index, matched 92 (4.5%) additional cases.

Table 1 shows the characteristics of breast cancer cases by presence versus absence on the sampling frame. Reference groups were chosen to represent commonly accepted groups with decreased risk of breast cancer. Case SPRs ranged from 0.71 to 1.12. Breast cancer cases found on the WDOT sampling frame were more likely to be older than 40 (SPR, 1.12) than breast cancer cases under the age of 40. A greater percentage of cases who were married, widowed, or single and never married were found on the WDOT master file than in cases who were divorced, separated, or living with a partner. Compared with cases with parity of 3 or more, all other cases were less likely to be found on the WDOT master file. Each one-child decrease in parity significantly predicted coverage on the WDOT sampling frame with a SPR of 0.96 [95% confidence interval (CI), 0.94–0.98; data not shown]. Compared with 71% of cases reporting earnings of less than $50,000, 66% of cases

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

reporting incomes over $50,000 and 64% of cases who did not report income were linked with a record from the WDOT sampling frame (data not shown). In addition, cases reporting the use of hormone replacement therapy were less likely to be covered by the WDOT sampling frame. Oral contraceptive use, body mass index at time of diagnosis, smoking status, alcohol consumption, menopause status, age of menarche, and family history of cancer were not associated with coverage on the WDOT sampling frame (data not shown).

Discussion

Awareness of selection bias in specific studies and study design types is important for both researchers and policy makers in public health. In case–control studies, selection bias may be compounded as cases and controls are often taken from different sampling frames but assumed to represent the same study base. Hence, selection is associated with disease outcome, and if this selection is also associated with specific risk factors, bias may result.

This study used breast cancer cases from an interview-based case–control study to examine coverage error of the WDOT sampling frame used to select controls from the population. This approach assumes that factors associated with coverage are similar in both case and control groups. In this database, 31% of cases (N = 926) found in the WDOT master file renewed their driver's license after being diagnosed with cancer. For those cases, we were unable to determine their opt-out status before diagnosis. However, removing those 31% of cases did not change the factors determined to be associated with "opt out". In addition, to argue that the determinants of opting out of the WDOT sampling frame are the same in this population of cases as in the general population, one would have to assume that nonparticipants have the same determinants of opting out as the cases that took part in this study, and that these determinants have the same magnitude of association with opting out. While this may be a valid assumption, more research is needed on the cases that did not participate in the WWHS study to evaluate whether the opt-out determinants remain consistent.

This research indicates that breast cancer cases under 40 years of age are more likely to "opt out". Bias due to this finding in breast cancer research would be expected to be minimal with approximately 6% of all incident breast cancer cases occurring in women under the age of 40 (21). However, the finding may forecast a future deterioration of the coverage of this sampling frame if one assumes that the majority of respondents that have opted out will likely not opt back into the list sold by the state. Another option available to researchers would be to simply exclude those cases where bias is most likely. In this case that would be women under the age of 40. The potential for misinterpretation of results may be more serious than the loss of precision due to the smaller sample size. However, potentially informative data on these subgroups would be lost. Researchers could also use interaction effects to evaluate stratum specific estimates of bias related to opt out in these subgroups.

Of interest to breast cancer research were the findings that parity, income, and the use of hormone replacement therapy were associated with coverage on the sampling frame used to select controls. Women with higher parity, lower income, and never users of postmenopausal hormones are also less likely to develop breast cancer. Using a sampling frame that has fewer women with high parity, fewer women with high incomes, and fewer women who have ever used hormones to select controls may bias the results of a case–control study when those risk factors are included as primary exposures or as covariates. By assuming the same SPRs for cases and controls, we estimate control SPRs < 1.0 for greater parity, higher income, and use of hormone replacement therapy. Control SPRs < 1.0 would create observed ORs that are numerically higher than the truth (positive bias) when evaluating these factors in case–control studies that use driver's license master files to

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

identify controls. For example, although there is widespread agreement that postmenopausal estrogen–progesterone therapy is a risk factor for the development of breast cancer, an OR from a case–control study that uses the driver's license master file to identify controls that enabled potential controls to opt out due to privacy concerns would likely be overestimated. For conditions other than breast cancer, these factors (income, postmenopausal hormone use, and parity) are also likely associated with socioeconomic variables.

One simple option for reducing selection bias may be to exclude cases not on the sampling frame used to select the controls. It is often assumed that this exclusion ensures that the study bases for cases and controls are comparable. However, this option requires one of the assumptions that we made in our investigation: that factors responsible for opting out are the same between cases and controls. Selection bias will remain even after excluding cases that could not be approached to serve as controls if this assumption is not met. In addition, the exclusion of potentially up to 32% of cases would be wasteful in terms of statistical power and precision. Besides using SPRs to approximate the selection bias due to inadequate sampling frame coverage, a researcher could calculate the predicted probability of coverage for each control and use inverse probability weighting or propensity scores to adjust for selection bias (17, 20, 22, 23). This approach has been used in previous research to adjust for variations in nonresponse between cases and control in population-based case–control studies on cancer etiology (20). This would require the additional assumptions that the coverage model was properly specified without omitting relevant factors or interaction terms.

Investigators should evaluate the comparability of each sampling frame's study base when designing the study. When coverage of the study base differs between cases and controls, investigators should calculate the expected direction of bias indicated by the SPR for each exposure of interest and use one of the established correction methods to adjust results, when a new exposure of interests is evaluated. A previous study (24) investigated the association of study participation with molecular exposures and concluded that, although there was no association, further research is needed. It should be noted that for selection to create bias, association with a risk factor is a necessary, but not sufficient, condition for bias, as the SPR can still be equal to 1. For SPRs not to equal 1, participation also needs to be associated with disease outcome conditionally on the risk factor, which can occur due to different sampling frames. Bias is less likely in a case–control study nested in a cohort study as, barring selective drop out, disease outcome occurs postenrollment. A similar argument that opt out is likely unrelated to case–control status, as it mostly occurred before cancer diagnosis, is used in our estimation approach.

Also, this study assumed that the case sampling frame, the Wisconsin Cancer Reporting System, was 100% complete. This allows for a simplified calculation of SPRs based on setting α and β to 1. There appears to be regional variation in reporting due to privacy concerns of neighboring states (15). Additional research could focus on the effects of this variation.

This study and linkage procedures have some limitations. In addition to the assumptions discussed earlier, some errors in linkage between WWHS cases and the WDOT master list of licensed drivers may have occurred. Linkage procedures could partially explain the observed association between marital status and coverage. The first 3 linkage procedures used last name and address when merging the driver's license list with WWHS case data. Last name and/or address often change when marital status changes. However, these data are normally updated when a license is renewed (every 7 years in Wisconsin). Also, linkages focusing on date of birth, aided by various weighting schemes, were evaluated to reduce misclassification. The WWHS parent study obtained a master list of licensed drivers in 2004

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

and 2006. However, due to cost constraints only the 2006 master file was prepared for linkage. Cases that were interviewed in 2004 may have moved away, passed away or had a disease progression that would explain absence from the 2006 master list of licensed drivers. However, the 2-year emigration and mortality rates are low for these women. Additional linkage errors probably resulted in nondifferential misclassification bias, resulting in attenuated SPRs.

Our results indicate the potential for selection bias due to differential opt out between various demographic and behavioral subgroups of controls. The potential for bias due to inadequate coverage of the study base will increase if more individuals opt out of inclusion in common sampling frames used in cancer research. All current control ascertainment schemes, including random digit dialing, have coverage issues, and as response rates and participation rates decline, understanding the effects of sampling frames that do not fully enumerate the study base will be critical to establishing the validity of study results.

Acknowledgments

The authors thank Leonelo Bautista, Nora Cate Schaeffer, Paul Peppard, John Hampton, Julie McGregor, and Laura Stephenson for study assistance provided.

Grant Support

This work was supported by the NIH grants CA47147, CA67264, and CD000712.

References

1. Choi BC, Pak AW. Understanding and minimizing epidemiologic bias in public health research. Can J Public Health. 2005; 96:284–286. [PubMed: 16625797]
2. Wacholder S, Silverman DT, McLaughlin JK, Mandel JS. Selection of controls in case-control studies. III. Design options. Am J Epidemiol. 1992; 135:1042–1050. [PubMed: 1595690]
3. Wacholder S, Silverman DT, McLaughlin JK, Mandel JS. Selection of controls in case-control studies. II. Types of controls. Am J Epidemiol. 1992; 135:1029–1041. [PubMed: 1595689]
4. Wacholder S, McLaughlin JK, Silverman DT, Mandel JS. Selection of controls in case-control studies. I. Principles. Am J Epidemiol. 1992; 135:1019–1028. [PubMed: 1595688]
5. Miettinen OS. The "case-control" study: valid selection of subjects. J Chronic Dis. 1985; 38:543–548. [PubMed: 4008595]
6. Rowland ML, Forthofer RN. Adjusting for nonresponse bias in a health examination survey. Public Health Rep. 1993; 108:380–386. [PubMed: 8497577]
7. Burgess AM Jr, Tierney JT. Bias due to nonresponse in a mail survey of Rhode Island physicians' smoking habits–1968. N Engl J Med. 1970; 282:908. [PubMed: 5434936]
8. Heilbrun LK, Nomura A, Stemmermann GN. The effects of nonresponse in a prospective study of cancer. Am J Epidemiol. 1982; 116:353–363. [PubMed: 7114044]
9. Beebe TJ, Talley NJ, Camilleri M, Jenkins SM, Anderson KJ, Locke GR III. The HIPAA authorization form and effects on survey response rates, nonresponse bias, and data quality: a randomized community study. Med Care. 2007; 45:959–965. [PubMed: 17890993]
10. Johnson TP, Holbrook AL, Ik Cho Y, Bossarte RM. Nonresponse error in injury-risk surveys. Am J Prev Med. 2006; 31:427–436. [PubMed: 17046415]
11. Walsh M, Trentham-Dietz A, Palta M. Availability of driver's license master-lists for use in government-sponsored public health research. Am J Epidemiol. 2011; 173:1414–1418. [PubMed: 21571870]
12. Lynch CF, Logsden-Sackett N, Edwards SL, Cantor KP. The driver's license list as a population-based sampling frame in Iowa. Am J Public Health. 1994; 84:469–472. [PubMed: 8129069]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

13. Adimora AA, Schoenbach VJ, Martinson FE, Stancil TR, Donaldson KH. Driver's license and voter registration lists as population-based sampling frames for rural African Americans. Ann Epidemiol. 2001; 11:385–388. [PubMed: 11454497]
14. Nichols HB, Trentham-Dietz A, Sprague BL, Hampton JM, Titus-Ernstoff L, Newcomb PA. Effects of birth order and maternal age on breast cancer risk: modification by whether women had been breastfed. Epidemiology. 2008; 19:417–423. [PubMed: 18379425]
15. Walsh, MC.; Stephanson, L.; Strickland, J.; Trentham-Dietz. Enhancing the completeness of Wisconsin Cancer Reporting System - The Border County Pilot Project. Vol. 2. Madison, WI: Surveillance Brief, University of Wisconsin Comprehensive Cancer Center; 2006. p. 1-4.
16. Horm, J. National Death Index Plus: Coded Causes of Death Supplement to the NDI User's Manual. Hyattsville, MD: Division of Vital Statistics National Center for Health Statistics; 1996. p. A-7–13
17. Palta, M. Quantitative methods in population health: extensions of ordinary regression. New York: Wiley; 2003.
18. Maclure M, Hankinson S. Analysis of selection bias in a case-control study of renal adenocarcinoma. Epidemiology. 1990; 1:441–447. [PubMed: 2090281]
19. Li Y, Graubard BI, DiGaetano R. Weighting methods for population-based case–control studies with complex sampling. J R Stat Soc Ser C. 2011; 60:21.
20. Colt JS, Schwartz K, Graubard BI, Davis F, Ruterbusch J, DiGaetano R, et al. Hypertension and risk of renal cell carcinoma among white and black Americans. Epidemiology. 2011; 22:797–804. [PubMed: 21881515]
21. Surveillance, Epidemiology, and End Results (SEER) Program. SEER*-Stat Database: Incidence - SEER 17 Regs Limited-Use +Katrina Impacted Louisiana Cases, Nov 2007 Sub (1973–2005 varying) - Linked to County Attributes - Total U.S., 1969–2005 Counties. National Cancer Institute, DCCPS, Surveillance Research Program, Cancer Statistics Branch. released April 2008, based on the Novembner 2007 submission, Contract No.: Document Number. Available from: www.seer.cancer.gov
22. Greenland S. Basic methods for sensitivity analysis of biases. Int J Epidemiol. 1996; 25:1107–1116. [PubMed: 9027513]
23. Kalton G, Piesse A. Survey research methods in evaluation and case-control studies. Stat Med. 2007; 26:1675–1687. [PubMed: 17278183]
24. Bhatti P, Sigurdson AJ, Wang SS, Chen J, Rothman N, Hartge P, et al. Genetic variation and willingness to participate in epidemiologic research: data from three studies. Cancer Epidemiology Biomarkers Prev. 2005; 14:2449–2453.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

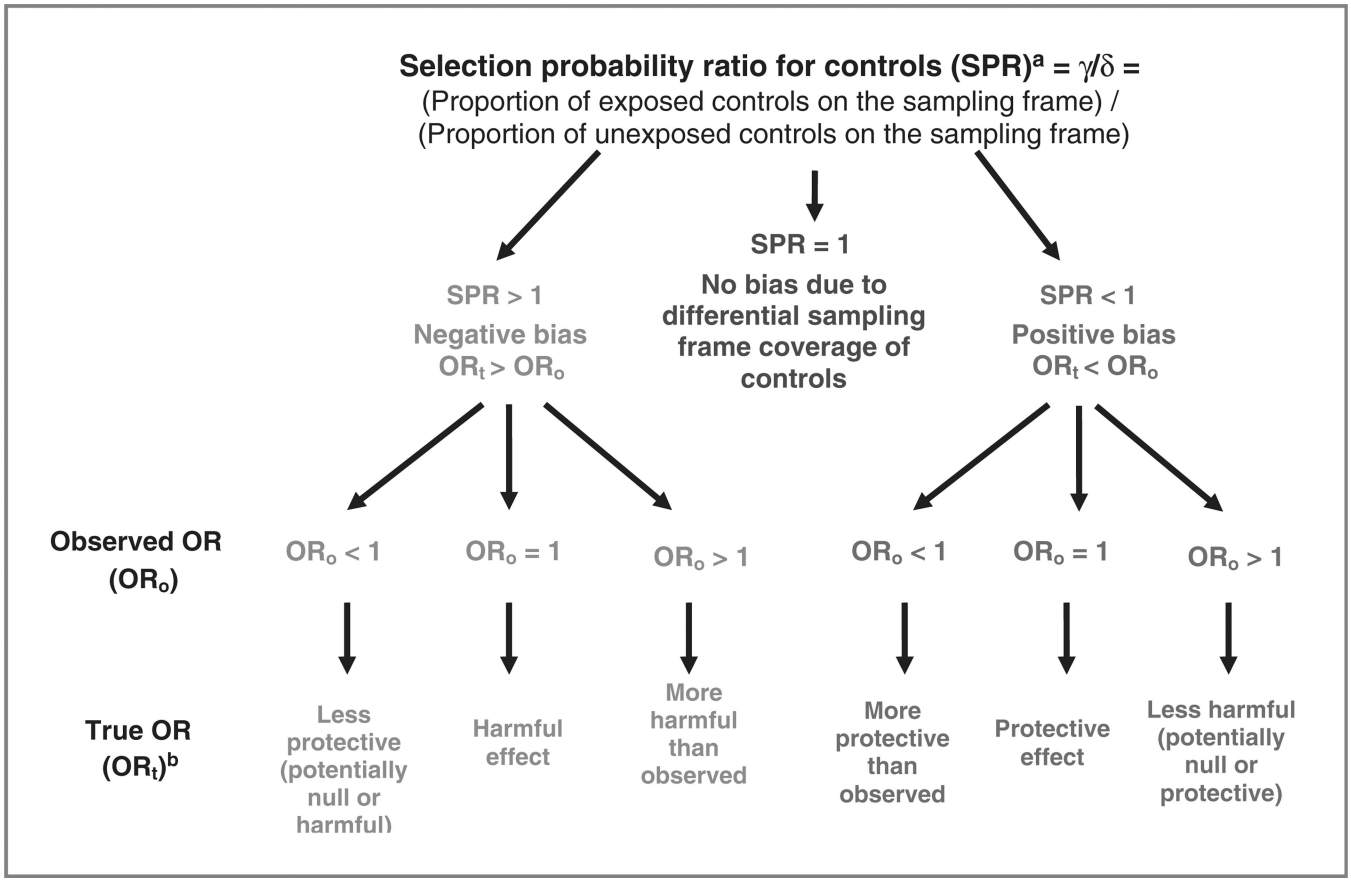

Figure 1.

Effects of differential sampling frame coverage of controls by exposure in a case–control study where the sampling frame for cases is assumed to have full case reporting. [a]In this study, under the assumption that opting out patterns are similar for cases and controls, we illustrate the likely magnitude of γ/δ via the corresponding ratio of probabilities of a case being on the drivers' license file used for controls. [b]True OR assuming no systematic or random error.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 1

SPRs for driver's license sampling frame coverage for breast cancer cases in the WWHS, 2004–2008

Characteristic at diagnosis	Percent of total cases[a] N = 2,988	SPR[b,c] (95% CI)
Age, y		
20–39	6.2	1 (reference)
40–69	93.8	1.12 (1.00–1.26)
Marital status at diagnosis		
Single, never married	5.4	1 (reference)
Married	77.6	0.90 (0.79–1.01)
Living with partner	2.2	0.71 (0.55–0.91)
Divorced, separated	9.1	0.83 (0.72–0.95)
Widowed	4.7	0.89 (0.76–1.04)
Income		
<$50,000	49.2	1 (reference)
≥$50,000	39.7	0.94 (0.89–1.00)
Missing	11.1	0.89 (0.81–0.98)
Postmenopausal hormone use		
Never	66.0	1 (reference)
Ever	33.2	0.93 (0.88–0.98)
Ever use of antidepressant medication use		
No	68.3	1 (reference)
Yes	31.66	0.95 (0.93–0.98)
Race		
White, non-Hispanic	95.1	1 (reference)
Other	3.4	1.06 (0.94–1.20)
Education		
College degree or more	40.6	1 (reference)
Some college	26.3	1.01 (0.94–1.08)
High school or less	33.1	1.02 (0.96–1.09)
Parity		
≥3	36.5	1 (reference)
2	36.7	0.94 (0.89–1.00)
1	11.7	0.86 (0.78–0.94)
0	14.3	0.87 (0.80–0.96)

[a] Not all categories sum to 100% due to missing values.

[b] ORs and 95% CIs were derived by fitting a generalized linear model with the log link function.

[c] Adjusted for parity, education, race, antidepressant medication use, hormone use, income, marital status at diagnosis, and age (over/under 40).