



Published in final edited form as:

*Cancer Causes Control*. 2009 July ; 20(5): 549–558. doi:10.1007/s10552-008-9263-1.

## Physical activity and premenopausal breast cancer: an examination of recall and selection bias

Sonia S. Maruti<sup>1,2,3,6</sup>, Walter C. Willett<sup>1,2,3</sup>, Diane Feskanich<sup>1</sup>, Beverly Levine<sup>5</sup>, Bernard Rosner<sup>1</sup>, and Graham A. Colditz<sup>1,2,4</sup>

<sup>1</sup>Channing Laboratory, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA

<sup>2</sup>Department of Epidemiology, Harvard School of Public Health, Boston, MA

<sup>3</sup>Department of Nutrition, Harvard School of Public Health, Boston, MA

<sup>4</sup>Department of Surgery, School of Medicine, Washington University, St. Louis, MO

<sup>5</sup>Department of Epidemiology, University of North Carolina at Greensboro, Greensboro, NC

<sup>6</sup>Department of Epidemiology, University of Washington; and Cancer Prevention Program, Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA

### Abstract

**Background**—Compared with cohort studies, case-control investigations have tended to report clearer protective associations for the relationship between physical activity and premenopausal breast cancer risk.

**Methods**—We conducted a case-control study within the Nurses' Health Study II cohort to examine whether recall or selection bias could explain the stronger protective associations. Self-reported total recreational physical activity during adulthood and over a woman's lifetime (ages 12 to current) were assessed in 1997 before diagnosis and again, one to seven years after breast cancer diagnosis among the same women.

**Results**—Eighty-seven percent of cases (417 of 479) and 82 percent of controls (390 of 474) responded. Selection bias was observed for activity during adulthood but not for activity over a woman's lifetime. Recall bias was not observed in the direction we expected: the odds ratios (ORs) for breast cancer comparing the highest versus lowest quintile of prospectively reported total activity were not significantly different than corresponding estimates from retrospective reports (e.g. lifetime activity: prospective OR=0.58, 95% CI: 0.37, 0.93 versus retrospective OR=0.80; 95% CI: 0.50, 1.29).

**Conclusion**—Recall or selection bias may not have accounted for protective associations among case-control investigations examining lifetime recreational physical activity and breast cancer. Selection bias related to recreational physical activity during adulthood and random error in measurement of physical activity remain concerns.

### Keywords

Physical Activity; Breast Cancer; Recall Bias; Selection Bias; Case-Control Study

## INTRODUCTION

In a previously published prospective analysis of the Nurses' Health Study cohort, we observed a 23% lower risk in premenopausal breast cancer for high levels of recreational physical activity over a woman's lifetime [1]. Although there is growing evidence for a possible benefit from physical activity for premenopausal breast cancer risk [2], the strength of the relationship has not been consistent in the epidemiological literature. Some case-control studies have reported even stronger protective associations [3–7] whereas most cohort studies have not observed associations of conclusive benefit [8, 9]. In a recent review, the average reduction in risk of overall breast cancer for the highest versus lowest categories of overall activity was 30% for cohort studies and 20% for case-control studies [2]. We thus investigated possible reasons for these differences. Because most studies have been case-control in design [2, 9], we assessed whether recall and selection biases [10] could have exaggerated the protective associations reported among case-control investigations. Understanding whether these biases could operate in the study of physical activity and breast cancer could help clarify the nature of this exposure-disease relationship.

Recall bias will arise if cases, because of their disease diagnosis, systematically over- or under-report their exposure compared with controls. For recall bias to result in exaggerated inverse associations, cases would have had to differentially underreport their activity compared with controls. This "differential" recall bias could occur if cases misremembered behaviors consistent with publicized exposure-disease associations or if lifestyle changes affected their memory of past behaviors. For example, some studies show that current diet can influence recall of past diet [11, 12]. Methodological studies have reported biased recall of tanning ability [13], fat intake [14, 15], family history of cancer [16], and induced abortion [17] among cancer patients, but there are no corresponding studies, to our knowledge, examining physical activity.

Selection bias is another potential concern [18], but has been less studied for cancer outcomes. Controls who participate in a case-control study may be more likely than nonresponders to be health conscious [10, 19, 20] and to be physically active. For selection bias to have exaggerated the protective associations in case-control studies of physical activity and breast cancer risk, women who chose to participate as controls would have to be more physically active than the entire group of women who were eligible to be controls.

To investigate whether recall and selection biases could be present in case-control studies, we conducted a case-control study nested within the Nurses' Health Study II (NHSII) cohort. We assessed recall bias by comparing assessments of recreational physical activity before and after breast cancer diagnosis among the same women. We investigated selection bias by comparing reports of responders and nonresponders.

## METHODS

### Nurses' Health Study Cohort

NHSII is an ongoing cohort study [21] established in 1989 when 116,608 female registered nurses aged 25 to 42 years completed a self-administered questionnaire about risk factors for chronic disease and medical history. Biennially, women are sent questionnaires to update this information and to inquire about additional risk factors. This study was approved by the Human Subjects Committee at Brigham and Women's Hospital in Boston, Massachusetts.

### Cases and controls

The source population consisted of 85,793 NHSII women who were premenopausal in 1997 (when a detailed assessment of physical activity was sent) with no previous report of cancer

except non-melanoma skin cancer (Figure 1). Breast cancer cases were initially identified among women who reported this diagnosis after 1997 on the subsequent NHSII questionnaires. Once permission to obtain medical records was granted, study physicians, blinded to the participant's exposure status, reviewed medical records and pathology reports to confirm self-reported diagnoses. Cases were defined as women who were diagnosed with invasive breast cancer from the return of the June 1997 questionnaire to June 2003, were premenopausal at diagnosis date, and had no previous report of cancer except non-melanoma skin cancer (n=604). Only incident cases were included.

Controls were randomly selected from the NHSII cohort and risk-set matched one-to-one with cases by year of birth. All controls (n=604) were premenopausal and had not reported a diagnosis of breast cancer or other cancer before or during the two-year questionnaire cycle when their matched cases was diagnosed.

The same cases and controls received the retrospective questionnaire. The retrospective questionnaire was mailed to new cases and matched controls in 2000 and also, to subsequently diagnosed cases identified up until 2004 and their matched controls. In the analyses, we excluded women who did not report any prospective activity data (100 cases and 97 controls) and those who did not report prospective activity between ages 12 to 22 years (20 cases and 27 controls), which were our main measures of interest in calculating lifetime activity. We also excluded 5 cases and 6 controls with outlying [22] activity data (1 percent of participants). Among these eligible women, 417 (87 percent) of 479 eligible cases and 390 (82 percent) of 474 eligible controls responded to the retrospective questionnaire and were available for this case-control analysis. The distribution of breast cancer risk factors was similar between those considered eligible and the 604 case-control pairs who were originally sampled; furthermore, response rates among eligible women were similar to that of the originally sampled cases (85% response) and controls (82% response). The median interval between date of diagnosis and receipt of the retrospective activity questionnaire among cases was 4 years; range 1 to 7 years. The mean age of case and control responders in 1997 was 43 years (range: 33 to 51 years).

### Prospective and Retrospective Assessments of Recreational Physical Activity

Prospective data were collected on the 1989 and 1997 NHSII questionnaires (available online [23]). Physical activity during adulthood was obtained when participants reported the hours per week (h/wk) they engaged in jogging, running, bicycling (including stationary machine), racquet sports, swimming laps, walking or hiking outdoors, calisthenics/aerobics, and other aerobic activity in the year prior to the 1989 or 1997 questionnaire. In 1997, participants also reported their physical activity during five age periods: grades 7–8 (junior high, ages 12–13), grades 9–12 (high school, ages 14–17), ages 18–22 (college), ages 23–29, and ages 30–34. For each age period, participants indicated the average h/wk they engaged in strenuous activity (e.g. running, aerobics, swimming laps), moderate activity (e.g. hiking, walking for exercise, casual cycling, yard work), walking to and from school/work, and television (TV) watching - our measure of inactivity. These activity data (excluding TV watching) were used in our calculation of total recreational activity over one's lifetime. We used these prospectively collected data as the "gold standard" for comparisons.

The reproducibility and validity of these prospective reports were good [24, 25]. Recalled activity for ages 12–22 had high 4-year reproducibility in a subgroup of 160 NHSII participants (average correlation  $r = 0.76$  for strenuous,  $r = 0.70$  for strenuous plus moderate, and  $r = 0.64$  for total physical activity) [24]. Also, our measure of physical activity during adulthood performed well when comparing reports from a questionnaire inquiring about activity in the previous year with recalls of past-week activity ( $r = 0.79$ ) and separately, with four seven-day activity diaries ( $r = 0.62$ ) among 149 representative NHSII participants [25].

Moreover, self-reported physical activity using a similar questionnaire was well correlated with lowered resting pulse ( $r = -0.45$ ) in men [26] and maximal oxygen consumption ( $r = 0.54$ ) in women [27].

Retrospective data were obtained starting in 2000 when the selected cases and controls completed questions identical to those in the prospective questionnaires about their physical activity during ages 12–34 and their past-year activity in 1989 and 1997. The cover letter stated that this study was about “better understanding the relationship between physical activity and a variety of women’s health outcomes, such as breast cancer and other diseases”. One reminder was sent.

### **Total Recreational Physical Activity During Adulthood and Lifetime**

We analyzed total recreational activity during adulthood and during a woman’s lifetime, because this allowed us to focus our investigation of methodological bias on a few key classifications of activity that other investigations have also reported. We have previously reported on breast cancer risk by different intensities and age-periods of physical activity [1].

To estimate total recreational activity, we first assigned each activity a metabolic equivalent value (MET) [25] based on the Compendium of Physical Activities [28]. For the five age-specific periods (ages 12–34), we assigned reports of strenuous, moderate, and walking activities MET values of 7.0, 4.5, and 3.0 respectively; the MET values were based on Center for Disease Control (CDC) designations of intensity categories [29]. Total recreational activity, expressed in MET-hours per week (MET- h/wk), was computed by multiplying the h/wk of each reported activity by their respective MET score, and summed the values. Total activity during adulthood was computed as a weighted sum of all reported activities (the weight being the MET value) on the 1997 questionnaire and was expressed in MET- h/wk. Because some studies have reported the average of repeated measures of activity, we also estimated the mean h/wk of total recreational activity during adulthood from both 1989 and 1997 assessments.

To obtain mean lifetime activity, we first used linear interpolation to estimate yearly adult activity between the last life period report for ages 30–34 and the past-year assessment in 1997. For example, in the case of a woman who was 40 in 1997, linear interpolation was used to estimate her activity for each age between 34 and 40, assuming that activity changed at a constant rate. We then estimated mean lifetime physical activity by averaging activity from age 12 to the participant’s current age. For example, lifetime total activity for a 50-year-old woman was estimated by summing her annual total activity at each age, from 12 to 50, and dividing this value by 39. TV watching, our measure of inactivity, was reported for ages 12 to 34 and was analyzed in h/wk.

### **Statistical analyses**

We investigated recall and selection bias in several ways. First, we assessed selection bias by comparing means of prospective activity reports of responders and nonresponders of the retrospective (e.g., case-control) questionnaire. Because activity values were skewed, we used the nonparametric Wilcoxon rank-sum test to determine the statistical significance of this comparison. We evaluated recall bias by comparing means of prospective versus retrospective reporting of physical activity among the same women. A Wilcoxon sign-rank test was used to determine statistical significance of the prospective versus retrospective reporting. Moreover, a Wilcoxon rank-sum test was used to compare the difference in prospective minus retrospective reporting of cases versus controls (differential recall bias).

To further examine recall bias, we used unconditional logistic regression models to obtain the prospective and retrospective odds ratios (OR) of breast cancer risk and 95% confidence intervals (CI). Unconditional logistic regression was used to analyze all case and control responders rather than only matched pairs. We categorized physical activity into quintiles according to the prospectively assessed exposure distribution of the control responders.

The odds ratios were adjusted for the following matching and breast cancer risk factors: age (years), whether in a previous blood study (no, yes), childhood body shape at ages 5 and 10 (as assessed by pictograms), duration and recency of oral contraceptive use (never, past < 4 yrs, past 4 yrs, current < 4 yrs, and current 4 yrs), mother or sister with breast cancer (no, yes), history of benign breast disease (no, yes), cross-classification of parity and age at first birth (afb) (nulliparous; parity 1–2, afb <25; parity 1–2, afb 25–29; parity 1–2, afb 30; parity 3, afb <25; parity 3, afb 25–29; parity 3, afb 30), current alcohol consumption (none, >0.0–1.4 g/day, 1.5–4.9 g/day, 5.0–9.9 g/day, 10 g/day), and adult height (inches).

To focus our evaluation of recall bias on physical activity, we included only covariates obtained from the prospectively collected NHSII questionnaires. Because controlling individually for body mass index (BMI) and age at menarche did not alter the odds ratios, and since these covariates may be intermediates[30] in the causal pathway between activity and breast cancer, we did not include them in our core models. We performed a test for linear trend by modeling the exposure as a continuous variable (outliers were excluded). Because of previous findings of biased recall of other exposures by methodological studies, we mainly focused our investigation on recall bias. A  $P < 0.05$  was considered statistically significant. All the analyses were performed with SAS version 9.0 (SAS Institute, Cary, NC).

## RESULTS

Table 1 describes age-adjusted participants' characteristics according to case and responder status. The study response rate was high, with more cases (87 percent) responding than controls (82 percent). In general, both nonresponding controls and cases had fewer breast cancer risk factors than their responding counterparts; however, this may have been due to chance because there were few nonresponders. As expected, cases were more likely than corresponding controls to have breast cancer risk factors.

We formally examined selection bias by comparing the prospective physical activity reports of controls who responded with those of nonresponding controls (Table 2). For selection bias to exaggerate protective associations, controls who agreed to participate would have had to be more physically active than the general pool of controls. For 1989 and 1997 averaged adult activity in controls, lower levels among nonresponders (14.2 MET-h/wk) versus responders (17.7 MET-h/wk) produced negative percent differences (–20%), suggesting selection bias for activity during adulthood ( $P=0.03$ ). Moreover, nonresponders, in addition to being less active, had greater levels of TV watching than responding controls (Table 1). No significant difference was observed in lifetime physical activity by response status in cases and controls. Physical activity levels among responding versus nonresponding cases were not appreciably different. Findings were similar when we further examined selection bias by restricting these analyses to those who completed the retrospective questionnaire before a reminder was sent (i.e., “first responders”).

Next, we examined recall bias by comparing the prospective versus retrospective reports according to case status (Table 3). For recall bias to exaggerate protective associations, cases would have had to differentially underreport their activity (retrospective lower than the

prospective report), leading to larger negative percent differences compared with controls. Contrary to expectation, cases tended to overreport their activity during adulthood (positive percent differences, column 3) more than controls. This was statistically significant for averaged 1989 and 1997 activity during adulthood ( $P=0.003$ ). Meanwhile, for lifetime activity, *both* cases and controls underreported their activity, with slightly larger negative percent differences among controls (percent difference in column 3: for cases,  $-7.8\%$ ; for controls,  $-12.9\%$ ); however, the difference in reporting between cases and controls was nonsignificant. Further, we did not find evidence for “differential” recall bias when examining means stratified by time interval between case diagnosis and retrospective questionnaire administration ( $< 2$  years versus  $>2$  years, data not shown).

Table 4 shows the multivariate odds ratios for breast cancer modeling quintiles of reported activity prospectively and, in separate regression models, retrospectively. Since this analysis was among responders we were only able to assess recall bias. We formally tested whether prospective and retrospective estimates were significantly different from each other by first entering the prospectively and retrospectively reported activities together in the same regression model as continuous terms and then evaluating the significance of the difference between their betas. Although most of the retrospective estimates were less strongly inverse than the prospective estimates, formal testing revealed no significant differences between the prospective and retrospective estimates. However, the prospective and retrospective reports of total activity were highly correlated ( $r > 0.60$ ) and this may have limited the ability to evaluate whether these estimates were different.

## DISCUSSION

To investigate recall and selection bias, we conducted a case-control study nested within the Nurses’ Health Study II (NHSII) cohort, comparing reports of recreational physical activity before and after breast cancer diagnosis among the same women (recall bias) and also, activity assessments of responders and nonresponders (selection bias). We found no evidence for “differential” recall bias to account for stronger protective associations reported in the case-control literature. However, we did observe evidence for selection bias, specifically, that participation among controls was associated with activity during adulthood.

At least 13 cohort studies [1, 31–42] and 21 case-control studies [3–7, 43–58] have reported estimates for recreational physical activity and premenopausal breast cancer risk (literature reviewed in [2, 8, 9]). When we initiated data collection for this nested case-control investigation in 2000, case-control studies reported a stronger protective association for premenopausal breast cancer than cohort studies. While our recent NHSII cohort investigation [1] suggests a protective association, we wanted to examine the issue of methodological bias, because results regarding recreational activity from cohort studies remain inconclusive.

For recreational activity during adulthood, most cohort and case-control investigations have not observed significant associations for breast cancer risk, potentially because women have not engaged in sufficiently high levels of activity. For activity during adolescence or over a lifetime, more case-control [3–7, 46, 49] than cohort [1, 41] investigations have reported statistically significant inverse associations. If the results of our current methodological study are generalizable, recall or selection bias is unlikely to account for the more consistently inverse associations among these case-control investigations.

Measurement error may be one potential reason for the difference between the case-control versus cohort results. Since, case-control studies have typically have used more detailed measures of physical activity than cohort studies, they may have suffered from less

measurement errors random error can attenuate associations. In this methodological study, with more detailed data on physical activity, we came close to the observed RRs of other case control studies. Still, our questionnaire asked about physical activity during certain age periods rather than ranges that were specific to individuals as have been used in some case-control studies of lifetime activity.

In our earlier, prospective analysis of the whole NHSII cohort (using the same physical activity questions) [1], we observed an inverse association between physical activity during women's lifetime (ages 12 to current age) and risk of breast cancer that is compatible with, although slightly weaker than, what we found prospectively in the subgroup included in this case-control analysis. Such results suggest that lifetime activity may be of particular benefit for premenopausal breast cancer risk.

Interestingly, in this methodological investigation, we observed that on the retrospective questionnaire, cases overreported their physical activity as adults compared with controls. Such overreporting of adult physical activity may have occurred if cases, as a consequence of their diagnoses, increased their physical activity, and this lifestyle change systematically influenced reporting of activity in the recent past before diagnosis. While some studies suggest that physical activity declines after diagnosis [59], one investigation [60], reported that patients were more physically active after diagnosis.

Our study has several limitations. Our investigation differed in some ways from a typical population-based case-control study. First, our average of 4 years between diagnosis and retrospective physical activity assessment does not normally reflect what happens in a case-control study in which rapid case ascertainment and interview are conducted. However, we have no reason to think, nor data to indicate, that women recently diagnosed with breast cancer are more likely than women diagnosed with breast cancer more than 2 years in the past to recall past exposures in a biased way. In post-hoc analyses, we did not find any evidence for greater "differential" recall bias when we stratified results by a  $\leq 2$  or  $>2$  year interval. Second, selection bias caused by choosing inappropriate controls was minimized, and thus, our assessment of this bias may be conservative, because controls were randomly sampled from the underlying enumerated cohort. Additionally, response rates of controls may have been higher than seen in typical case-control studies, potentially because all of our eligible case-control participants had to have responded, at the minimum, to the 1989 (baseline) and 1997 (age-specific physical activity) cohort questionnaires; thus, our findings of selection bias may be conservative. Some case-control studies that have observed apparent protective associations between physical activity and premenopausal breast cancer have reported low response rates ( $< 75$  percent) among controls, [44, 55, 61]. With regards to our physical activity measures, our focus was on mostly recreational activities (with questions on yard work and walking to and from work), and we did not report on occupational or indoor household activity. Lastly, the time span between retrospective questionnaire assessment and time of interest (e.g. years 1989, 1997, specific age periods) was longer than that for the prospective questionnaire. This would introduce more random measurement error in the retrospective than prospective activity assessment, as reflected by the weaker retrospective versus prospective odds ratios for risk of breast cancer (Table 4).

This study also has several strengths. We assessed activity before and after breast cancer diagnosis among the same women, using the same questions. Moreover, we focused our evaluation of methodological bias on physical activity and minimized residual confounding by utilizing covariate information from prospectively collected data. While several cancer studies have investigated recall and/or selection bias for diet [14, 15, 62–66] and other exposures [13, 16, 17], this is the first study, to our knowledge, of these biases for physical activity. In terms of generalizability, these results are applicable to Caucasian women.

Although participants were registered nurses at the initiation of the NHSII study, previous exposure-disease relations in the NHSII cohort have been confirmed in other study groups suggesting that our findings are generalizable on a population-level.

In summary, our findings, if generalizable to other study populations, suggest that recall and selection biases may not explain the stronger protective associations for adolescent/lifetime activity observed in some population-based case-control studies. Our data, however, do add to concerns regarding the possibility of selection bias due to nonparticipation of less physically active controls which could distort associations among case-control studies particularly if participation rates are low. This study also emphasizes the need for detailed exposure assessments to minimize random error.

## Acknowledgments

**Sources of Financial Support:** National Institutes of Health, National Cancer Institute (grant CA50385 and R25 CA98566 to S.S.M.). G.A.C was supported by an American Cancer Society Clinical Research Professorship.

We thank the NHSII women for their participation in this investigation. We are also grateful to Karen Corsano and Gary Chase for their assistance on the technical and logistical aspects of this study.

## References

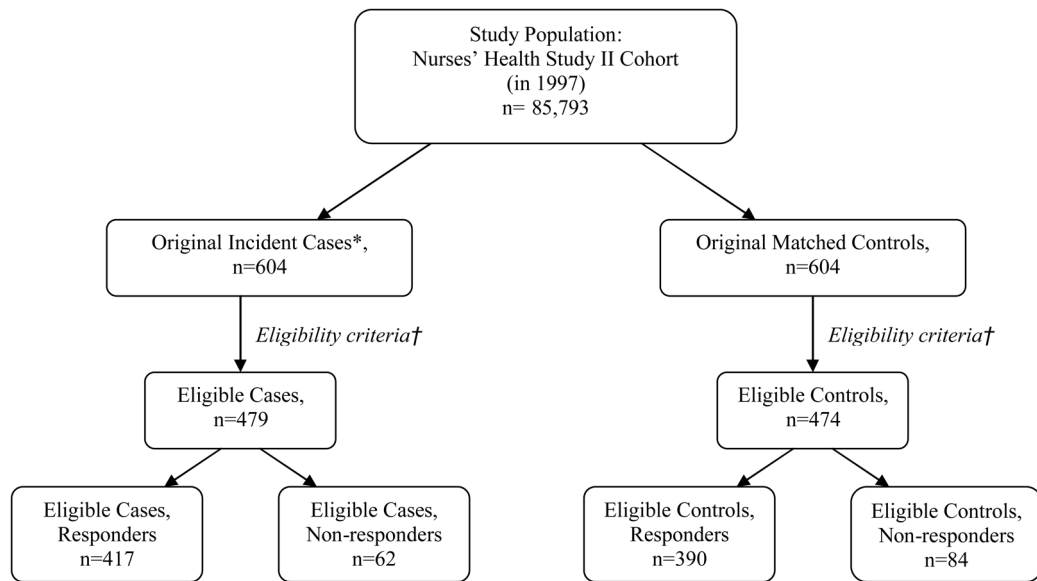
1. Maruti SS, Willett WC, Feskanich D, Rosner B, Colditz GA. A prospective study of age-specific physical activity and premenopausal breast cancer. *Journal of the National Cancer Institute*. 2008 May 21; 100(10):728–37. [PubMed: 18477801]
2. Friedenreich CM, Cust AE. Physical activity and breast cancer risk: impact of timing, type and dose of activity and population subgroup effects. *British journal of sports medicine*. 2008 Aug; 42(8): 636–47. [PubMed: 18487249]
3. Bernstein L, Henderson BE, Hanisch R, Sullivan-Halley J, Ross RK. Physical exercise and reduced risk of breast cancer in young women. *Journal of the National Cancer Institute*. 1994 Sep 21; 86(18):1403–8. [PubMed: 8072034]
4. Yang D, Bernstein L, Wu AH. Physical activity and breast cancer risk among Asian-American women in Los Angeles: a case-control study. *Cancer*. 2003 May 15; 97(10):2565–75. [PubMed: 12733156]
5. Adams-Campbell LL, Rosenberg L, Rao RS, Palmer JR. Strenuous physical activity and breast cancer risk in African-American women. *J Natl Med Assoc*. 2001 Jul-Aug;93(7–8):267–75. [PubMed: 11491277]
6. Dorn J, Vena J, Brasure J, Freudenheim J, Graham S. Lifetime physical activity and breast cancer risk in pre- and postmenopausal women. *Med Sci Sports Exerc*. 2003 Feb; 35(2):278–85. [PubMed: 12569217]
7. John EM, Horn-Ross PL, Koo J. Lifetime physical activity and breast cancer risk in a multiethnic population: the San Francisco Bay area breast cancer study. *Cancer Epidemiol Biomarkers Prev*. 2003 Nov; 12(11 Pt 1):1143–52. [PubMed: 14652273]
8. World Cancer Research Fund/American Institute for Cancer Research. *Food, nutrition, physical activity, and the prevention of cancer: a global perspective*. Washington, D.C: AICR; 2007.
9. Monninkhof EM, Elias SG, Vlems FA, van der Tweel I, Schuit AJ, Voskuil DW, et al. Physical activity and breast cancer: a systematic review. *Epidemiology*. 2007 Jan; 18(1):137–57. [PubMed: 17130685]
10. Austin H, Hill HA, Flanders WD, Greenberg RS. Limitations in the application of case-control methodology. *Epidemiol Rev*. 1994; 16(1):65–76. [PubMed: 7925729]
11. Rohan TE, Potter JD. Retrospective assessment of dietary intake. *Am J Epidemiol*. 1984 Dec; 120(6):876–87. [PubMed: 6507428]
12. Wu ML, Whittemore AS, Jung DL. Errors in reported dietary intakes. II. Long-term recall. *Am J Epidemiol*. 1988 Nov; 128(5):1137–45. [PubMed: 3189287]



13. Weinstock MA, Colditz GA, Willett WC, Stampfer MJ, Rosner B, Speizer FE. Recall (report) bias and reliability in the retrospective assessment of melanoma risk. *Am J Epidemiol.* 1991 Feb 1; 133(3):240–5. [PubMed: 2000841]
14. Giovannucci E, Stampfer MJ, Colditz GA, Manson JE, Rosner BA, Longnecker M, et al. A comparison of prospective and retrospective assessments of diet in the study of breast cancer. *Am J Epidemiol.* 1993 Mar 1; 137(5):502–11. [PubMed: 8465802]
15. Wilkens LR, Hankin JH, Yoshizawa CN, Kolonel LN, Lee J. Comparison of long-term dietary recall between cancer cases and noncases. *Am J Epidemiol.* 1992 Oct 1; 136(7):825–35. [PubMed: 1442748]
16. Floderus B, Barlow L, Mack TM. Recall bias in subjective reports of familial cancer. *Epidemiology.* 1990 Jul; 1(4):318–21. [PubMed: 2083311]
17. Rookus MA, van Leeuwen FE. Induced abortion and risk for breast cancer: reporting (recall) bias in a Dutch case-control study. *J Natl Cancer Inst.* 1996 Dec 4; 88(23):1759–64. [PubMed: 8944006]
18. Rothman, KJ.; Greenland, S. *Modern Epidemiology.* 2. Lippincott Williams & Wilkins; 1998.
19. Criqui MH, Austin M, Barrett-Connor E. The effect of non-response on risk ratios in a cardiovascular disease study. *J Chronic Dis.* 1979; 32(9–10):633–8. [PubMed: 489703]
20. Criqui MH, Barrett-Connor E, Austin M. Differences between respondents and non-respondents in a population-based cardiovascular disease study. *Am J Epidemiol.* 1978 Nov; 108(5):367–72. [PubMed: 727205]
21. Colditz GA, Hankinson SE. The Nurses' Health Study: lifestyle and health among women. *Nat Rev Cancer.* 2005 May; 5(5):388–96. [PubMed: 15864280]
22. Rosner B. Percentage points for a generalized ESD many-outlier procedure. *Technometrics.* 1983; 25(2):165–72.
23. Nurses' Health Study Website: Questionnaires. Channing Laboratory, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School; Boston, MA: 2007. (<http://www.channing.harvard.edu/nhs/questionnaires/index.shtml>). [Website] [cited 2007 June. 1]; Available from
24. Baer HJ, Schnitt SJ, Connolly JL, Byrne C, Willett WC, Rosner B, et al. Early life factors and incidence of proliferative benign breast disease. *Cancer Epidemiol Biomarkers Prev.* 2005 Dec; 14(12):2889–97. [PubMed: 16365006]
25. Wolf AM, Hunter DJ, Colditz GA, Manson JE, Stampfer MJ, Corsano KA, et al. Reproducibility and validity of a self-administered physical activity questionnaire. *Int J Epidemiol.* 1994 Oct; 23(5):991–9. [PubMed: 7860180]
26. Chasan-Taber S, Rimm EB, Stampfer MJ, Spiegelman D, Colditz GA, Giovannucci E, et al. Reproducibility and validity of a self-administered physical activity questionnaire for male health professionals. *Epidemiology.* 1996 Jan; 7(1):81–6. [PubMed: 8664406]
27. Jacobs DR Jr, Ainsworth BE, Hartman TJ, Leon AS. A simultaneous evaluation of 10 commonly used physical activity questionnaires. *Med Sci Sports Exerc.* 1993 Jan; 25(1):81–91. [PubMed: 8423759]
28. Ainsworth BE, Haskell WL, Leon AS, Jacobs DR Jr, Montoye HJ, Sallis JF, et al. Compendium of physical activities: classification of energy costs of human physical activities. *Med Sci Sports Exerc.* 1993 Jan; 25(1):71–80. [PubMed: 8292105]
29. Haskell WL, Lee IM, Pate RR, Powell KE, Blair SN, Franklin BA, et al. Physical activity and public health: updated recommendation for adults from the American College of Sports Medicine and the American Heart Association. *Med Sci Sports Exerc.* 2007 Aug; 39(8):1423–34. [PubMed: 17762377]
30. Bernstein L. Epidemiology of endocrine-related risk factors for breast cancer. *J Mammary Gland Biol Neoplasia.* 2002 Jan; 7(1):3–15. [PubMed: 12160084]
31. Colditz GA, Feskanich D, Chen WY, Hunter DJ, Willett WC. Physical activity and risk of breast cancer in premenopausal women. *Br J Cancer.* 2003 Sep 1; 89(5):847–51. [PubMed: 12942116]
32. Thune I, Brenn T, Lund E, Gaard M. Physical activity and the risk of breast cancer. *N Engl J Med.* 1997 May 1; 336(18):1269–75. [PubMed: 9113929]

33. Margolis KL, Mucci L, Braaten T, Kumle M, Trolle Lagerros Y, Adami HO, et al. Physical activity in different periods of life and the risk of breast cancer: the Norwegian-Swedish Women's Lifestyle and Health cohort study. *Cancer Epidemiol Biomarkers Prev.* 2005 Jan; 14(1):27–32. [PubMed: 15668472]
34. Rockhill B, Willett WC, Hunter DJ, Manson JE, Hankinson SE, Spiegelman D, et al. Physical activity and breast cancer risk in a cohort of young women. *Journal of the National Cancer Institute.* 1998 Aug 5; 90(15):1155–60. [PubMed: 9701365]
35. Breslow RA, Ballard-Barbash R, Munoz K, Graubard BI. Long-term recreational physical activity and breast cancer in the National Health and Nutrition Examination Survey I epidemiologic follow-up study. *Cancer Epidemiol Biomarkers Prev.* 2001 Jul; 10(7):805–8. [PubMed: 11440967]
36. Luoto R, Latikka P, Pukkala E, Hakulinen T, Vihko V. The effect of physical activity on breast cancer risk: a cohort study of 30,548 women. *Eur J Epidemiol.* 2000; 16(10):973–80. [PubMed: 11338130]
37. Albanes D, Blair A, Taylor PR. Physical activity and risk of cancer in the NHANES I population. *Am J Public Health.* 1989 Jun; 79(6):744–50. [PubMed: 2729471]
38. Lee SY, Kim MT, Kim SW, Song MS, Yoon SJ. Effect of lifetime lactation on breast cancer risk: a Korean women's cohort study. *International journal of cancer.* 2003 Jun 20; 105(3):390–3.
39. Sesso HD, Paffenbarger RS Jr, Lee IM. Physical activity and breast cancer risk in the College Alumni Health Study (United States). *Cancer Causes Control.* 1998 Aug; 9(4):433–9. [PubMed: 9794176]
40. Lahmann PH, Friedenreich C, Schuit AJ, Salvini S, Allen NE, Key TJ, et al. Physical activity and breast cancer risk: the European Prospective Investigation into Cancer and Nutrition. *Cancer Epidemiol Biomarkers Prev.* 2007 Jan; 16(1):36–42. [PubMed: 17179488]
41. Wyshak G, Frisch RE. Breast cancer among former college athletes compared to non-athletes: a 15-year follow-up. *Br J Cancer.* 2000 Feb; 82(3):726–30. [PubMed: 10682689]
42. Silvera SA, Jain M, Howe GR, Miller AB, Rohan TE. Energy balance and breast cancer risk: a prospective cohort study. *Breast cancer research and treatment.* 2006 May; 97(1):97–106. [PubMed: 16319973]
43. Gammon MD, Schoenberg JB, Britton JA, Kelsey JL, Coates RJ, Brogan D, et al. Recreational physical activity and breast cancer risk among women under age 45 years. *Am J Epidemiol.* 1998 Feb 1; 147(3):273–80. [PubMed: 9482502]
44. Steindorf K, Schmidt M, Kropp S, Chang-Claude J. Case-control study of physical activity and breast cancer risk among premenopausal women in Germany. *Am J Epidemiol.* 2003 Jan 15; 157(2):121–30. [PubMed: 12522019]
45. Friedenreich CM, Bryant HE, Courneya KS. Case-control study of lifetime physical activity and breast cancer risk. *Am J Epidemiol.* 2001 Aug 15; 154(4):336–47. [PubMed: 11495857]
46. Verloop J, Rookus MA, van der Kooy K, van Leeuwen FE. Physical activity and breast cancer risk in women aged 20–54 years. *Journal of the National Cancer Institute.* 2000 Jan 19; 92(2):128–35. [PubMed: 10639514]
47. Chen CL, White E, Malone KE, Daling JR. Leisure-time physical activity in relation to breast cancer among young women (Washington, United States). *Cancer Causes Control.* 1997 Jan; 8(1):77–84. [PubMed: 9051326]
48. Taioli E, Barone J, Wynder EL. A case-control study on breast cancer and body mass. The American Health Foundation--Division of Epidemiology. *Eur J Cancer.* 1995; 31A(5):723–8. [PubMed: 7640045]
49. Matthews CE, Shu XO, Jin F, Dai Q, Hebert JR, Ruan ZX, et al. Lifetime physical activity and breast cancer risk in the Shanghai Breast Cancer Study. *Br J Cancer.* 2001 Apr 6; 84(7):994–1001. [PubMed: 11286483]
50. Hu YH, Nagata C, Shimizu H, Kaneda N, Kashiki Y. Association of body mass index, physical activity, and reproductive histories with breast cancer: a case-control study in Gifu, Japan. *Breast cancer research and treatment.* 1997 Mar; 43(1):65–72. [PubMed: 9065600]
51. Slattery ML, Edwards S, Murtaugh MA, Sweeney C, Herrick J, Byers T, et al. Physical activity and breast cancer risk among women in the southwestern United States. *Ann Epidemiol.* 2007 May; 17(5):342–53. [PubMed: 17462544]

52. Coogan PF, Newcomb PA, Clapp RW, Trentham-Dietz A, Baron JA, Longnecker MP. Physical activity in usual occupation and risk of breast cancer (United States). *Cancer Causes Control*. 1997 Jul; 8(4):626–31. [PubMed: 9242479]
53. Friedenreich CM, Rohan TE. Physical activity and risk of breast cancer. *Eur J Cancer Prev*. 1995 Apr; 4(2):145–51. [PubMed: 7767240]
54. Kruk J, Aboul-Enein HY. Occupational physical activity and the risk of breast cancer. *Cancer Detect Prev*. 2003; 27(3):187–92. [PubMed: 12787725]
55. Gilliland FD, Li YF, Baumgartner K, Crumley D, Samet JM. Physical activity and breast cancer risk in hispanic and non-hispanic white women. *Am J Epidemiol*. 2001 Sep 1; 154(5):442–50. [PubMed: 11532786]
56. Hirose K, Hamajima N, Takezaki T, Miura S, Tajima K. Physical exercise reduces risk of breast cancer in Japanese women. *Cancer Sci*. 2003 Feb; 94(2):193–9. [PubMed: 12708496]
57. Sprague BL, Trentham-Dietz A, Newcomb PA, Titus-Ernstoff L, Hampton JM, Egan KM. Lifetime recreational and occupational physical activity and risk of in situ and invasive breast cancer. *Cancer Epidemiol Biomarkers Prev*. 2007 Feb; 16(2):236–43. [PubMed: 17301255]
58. Ueji M, Ueno E, Osei-Hyiaman D, Takahashi H, Kano K. Physical activity and the risk of breast cancer: a case-control study of Japanese women. *J Epidemiol*. 1998 Jun; 8(2):116–22. [PubMed: 9673081]
59. Blanchard CM, Denniston MM, Baker F, Ainsworth SR, Courneya KS, Hann DM, et al. Do adults change their lifestyle behaviors after a cancer diagnosis? *Am J Health Behav*. 2003 May-Jun; 27(3):246–56. [PubMed: 12751621]
60. Patterson RE, Neuhouser ML, Hedderson MM, Schwartz SM, Standish LJ, Bowen DJ. Changes in diet, physical activity, and supplement use among adults diagnosed with cancer. *J Am Diet Assoc*. 2003 Mar; 103(3):323–8. [PubMed: 12616253]
61. Friedenreich CM, Courneya KS, Bryant HE. Influence of physical activity in different age and life periods on the risk of breast cancer. *Epidemiology*. 2001 Nov; 12(6):604–12. [PubMed: 11679785]
62. Giovannucci E, Stampfer MJ, Colditz GA, Manson JE, Rosner BA, Longnecker MP, et al. Recall and selection bias in reporting past alcohol consumption among breast cancer cases. *Cancer Causes Control*. 1993 Sep; 4(5):441–8. [PubMed: 8218876]
63. Friedenreich CM, Howe GR, Miller AB. An investigation of recall bias in the reporting of past food intake among breast cancer cases and controls. *Ann Epidemiol*. 1991 Aug; 1(5):439–53. [PubMed: 1669524]
64. Friedenreich CM, Howe GR, Miller AB. Recall bias in the association of micronutrient intake and breast cancer. *J Clin Epidemiol*. 1993 Sep; 46(9):1009–17. [PubMed: 8263573]
65. Holmberg L, Ohlander EM, Byers T, Zack M, Wolk A, Bruce A, et al. A search for recall bias in a case-control study of diet and breast cancer. *Int J Epidemiol*. 1996 Apr; 25(2):235–44. [PubMed: 9119547]
66. Mannisto S, Pietinen P, Virtanen M, Kataja V, Uusitupa M. Diet and the risk of breast cancer in a case-control study: does the threat of disease have an influence on recall bias? *J Clin Epidemiol*. 1999 May; 52(5):429–39. [PubMed: 10360338]



**Fig. 1.** Flow chart of nested case-control study. Original incident cases (\*) were defined as women who were diagnosed with invasive breast cancer from June 1997 to June 2003, were premenopausal at date of diagnosis, and had no previous report of cancer except non-melanoma skin cancer. Cases and controls were excluded (†) if they did not have prospective physical activity data or if reported their physical activity values. were

Table 1

Characteristics of all eligible, responding, and nonresponding women in a case-control study nested within the Nurses' Health Study II cohort, United States, 1997 <sup>a</sup>

	All eligible		Responders		Nonresponders	
	Cases (n=479)	Controls (n=474)	Cases (n=417)	Controls (n=390)	Cases (n=62)	Controls (n=84)
Age (yrs)	43.4	43.3	43.5	43.4	43.1	42.9
Mother or sister with breast cancer (%)	15.4%	9.75%	16.0%	10.3%	10.7%	7.40%
History of benign breast disease (%)	23.7%	15.9%	24.6%	15.7%	18.0%	16.6%
Menarche 12 years (%)	24.6%	21.3%	24.8%	21.6%	23.9%	20.3%
Nulliparous (%)	20.9%	20.9%	22.0%	21.9%	11.6%	16.7%
Parity <sup>b</sup>	2.17	2.39	2.18	2.37	2.15	2.50
Age at first birth (yrs) <sup>b</sup>	27.5	26.6	27.5	26.5	27.5	26.7
Current oral contraceptive user (%)	8.18%	9.66%	8.45%	9.92%	7.12%	8.21%
Height (in)	65.2	64.8	65.2	64.8	65.2	64.7
Birthweight, 8.5 lbs (%)	13.0%	10.5%	13.8%	9.90%	8.24%	13.1%
Overweight, ages 5 and 10 (%)	3.37%	5.04%	3.16%	5.10%	5.22%	4.64%
BMI at age 18 (kg/m <sup>2</sup> )	21.0	21.2	20.9	21.2	21.6	21.3
Current BMI (kg/m <sup>2</sup> )	25.6	25.5	25.5	25.6	26.6	25.4
Alcohol intake, 10 g/day (%) <sup>c</sup>	11.4%	10.4%	12.4%	11.0%	4.75%	7.02%
Animal fat (% energy) <sup>c</sup>	17.1	17.0	17.0	17.1	18.6	16.6
Multivitamin use (%)	51.7%	54.1%	52.0%	54.8%	47.6%	50.1%
Current smoker (%)	12.1%	9.26%	11.7%	9.64%	14.1%	7.05%
Television watching from ages 12-34 (h/wk)	7.84	8.36	7.87	8.14	7.57	9.54

<sup>a</sup> All means and percentages refer to age-standardized prospective data collected in 1997, unless otherwise noted.

<sup>b</sup> Among parous women only.

<sup>c</sup> Refers to prospective data collected in 1995.

**Table 2**

Assessment of selection bias by comparing means of total physical activity among responders and nonresponders of a nested case-control study within the Nurses' Health Study II cohort <sup>a, b</sup>

	<b>Mean prospective total physical activity (MET-h/wk)</b>			
	<b>Column 1</b>	<b>Column 2</b>	<b>Column 3</b>	<b>Column 4</b>
	<b>All eligible women</b>	<b>Responders</b>	<b>Nonresponders</b>	<b>% Difference Selection bias <math>\left[ \frac{col3-col2}{col2} \right] * 100</math></b>
<b>Adult physical activity</b>				
<b>1997</b>				
Cases	11.6	11.6	11.3	-2.58%
Controls	13.8	14.0	13.2	-5.38%
<b>1989 &amp; 1997, average</b>				
Cases	14.9	14.9	15.1	1.28%
Controls	17.1	17.7	14.2	-20.0% <sup>c</sup>
<b>Lifetime physical activity</b>				
Cases	32.6	32.8	30.9	-5.70%
Controls	35.9	36.4	33.6	-7.69%

<sup>a</sup>Activity data was square-root transformed to improve normality for statistical tests. For interpretability, activity data are presented as back-transformed (squared) means. Percent difference were calculated using original, unrounded numbers.

<sup>b</sup>Numbers of cases and controls who were eligible, responders, and nonresponders are provided in the heading of Table 1.

<sup>c</sup>A Wilcoxon rank-sum test was conducted comparing reporting of responders versus nonresponders, for cases and controls separately; *P* is significant at the 0.05 level indicating a statistical difference between reporting of physical activity among responders versus nonresponders.

**Table 3**

Assessment of recall bias by comparing means for prospective and retrospective reporting of physical activity among responders to a nested case-control study within the Nurses' Health Study II cohort <sup>a, b</sup>

<u>Mean total physical activity (MET-h/wk), responders</u>			
	<u>Column 1</u>	<u>Column 2</u>	<u>Column 3</u>
	Prospective	Retrospective	% Difference (recall bias) $\left[ \frac{Col\ 2 - Col\ 1}{Col\ 2} \right] * 100$
<b>Adult physical activity</b>			
<b>1997</b>			
Cases	11.6	18.6	60.5% <sup>c</sup>
Controls	14.0	19.5	39.6% <sup>c</sup>
<b>1989 &amp; 1997, average</b>			
Cases	14.9	20.2	35.3% <sup>c,d</sup>
Controls	17.7	20.5	15.4% <sup>c</sup>
<b>Lifetime physical activity</b>			
Cases	32.8	30.3	-7.77% <sup>c</sup>
Controls	36.4	31.7	-12.9% <sup>c</sup>

<sup>a</sup>There were 417 case and 390 control responders.

<sup>b</sup>Activity data was square-root transformed to improve normality for statistical tests. For interpretability, activity data are presented as back-transformed (squared) means.

<sup>c</sup>Wilcoxon sign-rank test comparing the difference between prospective and retrospective reporting, for cases and separately, controls; *P* is significant at the 0.05 level indicating a statistical difference between prospective versus retrospective reporting.

<sup>d</sup>Wilcoxon rank-sum test comparing retrospective minus prospective reporting for cases versus controls; *P* is significant at the .05 level, indicating a statistical difference between case versus control reporting.

Assessment of recall bias by comparing multivariate breast cancer odds ratios of prospectively and retrospectively reported physical activity from 417 cases and 390 controls in a nested case-control study within the Nurses' Health Study II cohort, 1997–2003 <sup>a</sup>

**Table 4**

	<u>Quintiles of total physical activity (MET-h/wk) <sup>b</sup></u>					<i>P</i> for Trend <sup>c</sup>	
	1 (reference)	2	3	4	5		95% CI for Q5 vs. Q1
<b>Adult physical activity</b>							
<b>1997</b>							
Prospective	1.00	1.22	1.08	0.85	0.84	0.52,1.36	0.12
Retrospective	1.00	1.05	0.74	1.08	0.83	0.50,1.35	0.69
<b>1989 &amp; 1997, average</b>							
Prospective	1.00	0.80	0.61	0.64	0.67	0.43,1.06	0.11
Retrospective	1.00	1.26	1.03	0.80	1.04	0.66,1.64	0.80
<b>Lifetime physical activity</b>							
Prospective	1.00	1.08	0.80	0.48	0.58	0.37,0.93	0.009 <sup>c</sup>
Retrospective	1.00	0.85	0.60	0.62	0.80	0.50,1.29	0.38

<sup>a</sup>The odds ratios were adjusted for the following matching and breast cancer risk factors: age (years), whether in a previous blood study (no, yes), childhood body shape at ages 5 and 10 (as assessed by pictograms), duration and recency of oral contraceptive use (never, past < 4 yrs, current < 4 yrs, and current ≥ 4 yrs), mother or sister with breast cancer (no, yes), history of benign breast disease (no, yes), cross-classification of parity and age at first birth (afb) (nulliparous; parity 1–2, afb <25; parity 1–2, afb 25–29; parity 3, afb <25; parity 3, afb 25–29; parity 3, afb ≥ 30), current alcohol consumption (none, >0.0–1.4 g/day, 1.5–4.9 g/day, 5.0–9.9 g/day, 10 g/day), and adult height (inches).

<sup>b</sup>Quintiles were defined from the prospective reports of controls who responded to the study.

<sup>c</sup>Test of trend was calculated using activity modeled continuously; *P* for trend is statistically significant at the 0.05 level.