



Published in final edited form as:

*J Am Stat Assoc.* 2008 December 1; 103(484): 1392–1404. doi:10.1198/016214508000000706.

## Does Finasteride Affect the Severity of Prostate Cancer? A Causal Sensitivity Analysis

**Bryan E. Shepherd,**

Department of Biostatistics, Vanderbilt University, Nashville, TN, 37232, USA

**Mary W. Redman,** and

Fred Hutchinson Cancer Research Center, Seattle, WA, 98109, USA

**Donna P. Ankerst**

Department of Biostatistics, University of Munich, Munich, Germany

### Abstract

In 2003 Thompson and colleagues reported that daily use of finasteride reduced the prevalence of prostate cancer by 25% compared to placebo. These results were based on the double-blind randomized Prostate Cancer Prevention Trial (PCPT) which followed 18,882 men with no prior or current indications of prostate cancer annually for seven years. Enthusiasm for the risk reduction afforded by the chemopreventative agent and adoption of its use in clinical practice, however, was severely dampened by the additional finding in the trial of an increased absolute number of high-grade (Gleason score  $\geq 7$ ) cancers on the finasteride arm. The question arose as to whether this finding truly implied that finasteride increased the risk of more severe prostate cancer or was a study artifact due to a series of possible post-randomization selection biases, including differences among treatment arms in patient characteristics of cancer cases, differences in biopsy verification of cancer status due to increased sensitivity of prostate-specific antigen under finasteride, differential grading by biopsy due to prostate volume reduction by finasteride, and nonignorable drop-out. Via a causal inference approach implementing inverse probability weighted estimating equations, this analysis addresses the question of whether finasteride caused more severe prostate cancer by estimating the mean treatment difference in prostate cancer severity between finasteride and placebo for the principal stratum of participants who would have developed prostate cancer regardless of treatment assignment. We perform sensitivity analyses that sequentially adjust for the numerous potential post-randomization biases conjectured in the PCPT.

### Keywords

Causal inference; principal stratification; treatment effects; selection bias

## 1. INTRODUCTION

The Prostate Cancer Prevention Trial (PCPT) was a multi-center, double blind, randomized trial that studied the effect of finasteride on the period prevalence of prostate cancer in healthy men screened for 7 years (Thompson et al. 2003). The 18,882 men aged 55 years or older with no history or current indicators of prostate cancer (prostate-specific antigen (PSA)  $\leq 3.0$  nanograms per milliliter (ng/mL) and digital rectal exam (DRE) normal) were randomized to receive either 5 milligrams of finasteride per day or placebo and followed for 7 years. During annual follow-up, participants were referred for a prostate biopsy if their PSA exceeded a threshold or their DRE was abnormal (suspicious for cancer). In addition, all participants not diagnosed with prostate cancer during the study were instructed to undergo an end-of-study prostate biopsy at their seventh and final visit.

Of the 10168 men whose cancer status was known either by a positive mid-study biopsy or a study endpoint biopsy, prostate cancer was detected in 821 (16.6%) of 4951 men on the finasteride arm compared with 1194 (22.9%) of 5217 men on the placebo arm, suggesting that finasteride lowered the risk of prostate cancer ( $P < 0.001$ ). However, 299 (36.4%) of the 821 finasteride prostate cancer cases were more severe (Gleason score  $\geq 7$ ) compared to only 264 (22.1%) of 1194 placebo prostate cancer cases ( $P < 0.001$ ); see Table 1. Interpretation of the results is therefore challenging since the study suggested that finasteride reduced the overall risk of prostate cancer but accelerated growth of high-grade tumors (Scardino 2003).

While the apparent proportion of high-grade cancers among those diagnosed with cancer on finasteride is higher than that on placebo, this may not be an appropriate measure of finasteride's effect on disease severity. Those men diagnosed with cancer are a subset of those men initially randomized in the trial. As this subset was selected after randomization, there could be selection bias (Rosenbaum 1984). If the characteristics of men diagnosed with prostate cancer differ between treatment arms, then the apparent effect of finasteride on prostate cancer grade may be due to correlations between these differing characteristics and cancer grade, rather than the causal effect of finasteride. Additionally, using the number of cancers as the denominator (instead of the number of biopsies or the number randomized) ignores the possibility that finasteride prevented a large fraction of low-grade cases. An example of how post-randomization selection could influence results is shown in Figure 1.

To limit such potential selection bias, one could compare the prevalence of high-grade cancer among those who received a biopsy (as opposed to among those who were diagnosed with cancer). Of the 4951 men who had a biopsy in the finasteride arm, 6.0% had a high-grade tumor whereas 5.1% of the 5217 men with a biopsy in the placebo arm had a high-grade tumor, a difference that is borderline statistically significant;  $P = 0.03$ . Such an analysis may be important for public health purposes. However, it does not directly address the question of determining the effect of finasteride on cancer severity, but presents the combined effects of finasteride on cancer prevalence and on cancer severity among prevalent cases.

A relevant population for addressing the effect of finasteride on cancer severity is the subgroup of patients who would have developed cancer regardless of treatment, but whose treatment may have affected severity (Robins 1995; Rubin 2000; Frangakis and Rubin 2002). The potential outcomes framework (Neyman 1923; Rubin 1978) can be used to define this population. Specifically, as shown in Table 2, participants can be classified into four categories of paired potential outcomes (principal strata (Frangakis and Rubin 2002)) under finasteride and placebo: a participant could have never developed prostate cancer regardless of treatment assignment (stratum NN), a participant could have developed prostate cancer only if they received placebo (stratum CN), developed it only if they received finasteride (stratum NC), or developed prostate cancer regardless of treatment assignment (stratum CC). The total number of high-grade cancers on placebo comes from the number of high-grade cancers in strata CN and CC; the total number of high-grade cancers on finasteride comes from the total number of high-grade cancers in strata NC and CC. While discordance in the contribution of strata CN and NC to the total number of high grades is related to both the ability of finasteride to prevent prostate cancer and affect its severity; within stratum CC, differences in the number of high grade cancers between treatment arms is restricted solely to the effect of finasteride on cancer severity.

Hudgens, Hoering, and Self (2003), Gilbert, Bosch, and Hudgens (2003), Hudgens and Halloran (2006), Zhang and Rubin (2003), and Jemai (2005) have discussed methods, primarily in the context of vaccine trials, for assessing treatment effects on outcomes defined by a post-randomization event. In this manuscript, we first adapt these methods to estimate the effect of finasteride on prostate cancer severity among those who would have had biopsy-

detectable cancer regardless of treatment assignment. A key ingredient of our proposed implementation is a sensitivity analysis. In addition to uninformed sensitivity analyses that vary sensitivity parameters over their entire range, we highlight results based on elicited parameters from two subject matter experts. Next we extend these methods to accommodate two additional layers of potential bias in the PCPT: differential biopsy verification and differential cancer grading between finasteride and placebo.

Despite the PCPT protocol's specification that participants undergo an end-of-study biopsy, not all participants received a biopsy. This was expected as prostate biopsy is an invasive procedure with possible negative side effects (Goodman et al. 2004). As shown in Table 1, 65.0% of men randomized to the placebo arm received a biopsy compared to 62.2% on finasteride ( $P = 0.0002$ ). This difference is small with statistical significance driven by the large sample size. However, the assumption that biopsies were missing completely at random (MCAR) (Little and Rubin 2002) is unlikely for these data since one of the criteria for interim biopsies was PSA exceeding 4.0 ng/mL or an abnormal DRE (Baker 2000). A differential biopsy verification process between finasteride and placebo is also probable. Finasteride shrinks the volume of the prostate, so it approximately halves the PSA value. Therefore the actual PSA criterion for referral to biopsy on the finasteride arm was referral if an inflation factor (approximately 2.0)  $\times$  PSA exceeded 4.0 ng/mL (Thompson et al. 2003). Thompson et al. (2006, 2007) showed that the sensitivity of PSA was greatly enhanced on finasteride, and that shrinkage of the prostate gland on finasteride also affected the sensitivity of the DRE test. Since referral to biopsy was strictly mandated by protocol based on observed covariates PSA and DRE, the missing data process is more likely missing at random (MAR) (Little and Rubin 2002; Thompson et al. 2005). Our approach will be to assume MAR by incorporating the observed covariates PSA, DRE, family history of prostate cancer, age, race, and history of a prior negative biopsy (see Table 3) to modify the estimating equations using an inverse weighting approach similar to that proposed by Robins, Rotnitzky and Zhao (1995).

A second potential bias that is hypothesized to have driven the increased number of high-grade prostate cancer cases on the finasteride arm is due to differential biopsy grading (Lucia et al. 2007). The PCPT used 6-core biopsies which extract six biopsy tissues uniformly spaced across the prostate. Because finasteride shrinks the prostate volume, the 6-core biopsies covered a larger area of the prostate for cases in the finasteride arm and hence were probably more likely to detect high-grade prostate cancer than on the placebo arm. To investigate this hypothesis the PCPT performed a limited, blinded follow-up study of grade on prostatectomy for 531 PCPT participants (225 placebo, 306 finasteride) diagnosed with prostate cancer on biopsy during the study who were later treated by prostatectomy. In the placebo group the sensitivity of biopsy for high-grade detection was 45% (55 biopsy high-grades / 123 prostatectomy high-grades), compared to 66% on finasteride (76 biopsy high-grades / 115 prostatectomy high-grades), suggesting a substantial downward bias in detecting high-grade cancer on placebo relative to finasteride (Table 4). We will consider the impact of this differential biopsy grading by performing an analysis using cancer grades based on prostatectomy. Those men for whom we were able to obtain a prostatectomy tissue sample were not randomly selected from those who were diagnosed with prostate cancer; they tended to be younger, non African American, to have higher PSAs, worse DREs, and have higher grade on biopsy than those without a prostatectomy (Table 5). We use a similar inverse weighting procedure to accommodate this missing data mechanism, incorporating grade on biopsy as an additional covariate.

In section 2, we formulate our problem using a causal modeling framework. We propose three assumptions that identify the average causal effect of finasteride on severity of cancer among participants who would have developed biopsy-detectable cancer regardless of treatment assignment. In section 3, we perform sensitivity analyses investigating the implications of plausible violations to these assumptions. Sensitivity analyses are performed by systematically

relaxing assumptions, estimating sharp bounds and using sensitivity parameters to examine a gradation of assumption violations. Results from the PCPT are shown over sensitivity parameter ranges elicited from two subject matter experts. In section 4, we perform an analysis accounting for possible differential biopsy grading. And in section 5, we discuss our findings and the general analytic approach. Relevant estimating equations are found in the Appendix.

## 2. CAUSAL MODELING FRAMEWORK

### 2.1 Potential Outcomes and Observed Data

As previously mentioned, to define our target of interest we use a causal modeling framework which employs potential outcomes (Neyman 1923; Rubin 1978). Let  $Z = 0$  or  $1$  denote assignment to placebo or finasteride, respectively. Let  $S_i(z)$  be the indicator of biopsy-detectable prostate cancer if, possibly contrary to fact, subject  $i$  is assigned  $Z_i = z$ . (It should be noted that for brevity, throughout this paper we will often write “got cancer” or “developed cancer” instead of the more precise “had a biopsy-detectable prostate cancer.” The implications of any differential sensitivity of the 6-core biopsies used to detect prostate cancer in this trial will be discussed in section 5.) Similarly, let  $R_i(z)$  be the indicator that  $S_i(z)$  is known if  $Z_i = z$ ; consistent with the primary analysis of the PCPT,  $R_i(z) = 1$  if a participant has an end of study biopsy or a positive mid-study biopsy given  $Z_i = z$ . Let  $Y_i(z)$  be the indicator of high-grade prostate cancer (Gleason grade  $\geq 7$ ) if  $Z_i = z$ . (Gleason grades are subjective, ordinal scores which we have chosen to dichotomize at the traditional value to be consistent with the previous PCPT analyses and subsequent controversy.) In this setting, if a subject does not have prostate cancer, then the severity of their cancer does not exist. Specifically, we set  $Y_i(z) = *$  if  $S_i(z) = 0$ . Additionally, let  $X_i(z)$  be a vector of covariates if  $Z_i = z$ . Note that  $X_i(0) = X_i(1)$  for baseline covariates.

In order to link potential outcomes to observed data we assume that

$$(R_i(z), S_i(z), Y_i(z), X_i(z)) = (R_i, S_i, Y_i, X_i) \text{ if } Z_i = z, \quad (1)$$

where  $R_i$  is the observed indicator that prostate cancer status is known,  $S_i$  is the realized indicator of prostate cancer (missing if  $R_i = 0$ ),  $Y_i$  is the realized indicator of high-grade prostate cancer (missing if  $R_i = 0$  and not defined if  $S_i = 0$ ), and  $X_i$  are the observed covariates. This notation implicitly assumes that the potential outcomes of each trial participant are not influenced by the treatments of other participants, known as the Stable Unit Treatment Value Assumption (SUTVA) (Rubin 1978) or “no interaction between units” (Cox 1958). Let the observed data  $O_i = (Z_i, R_i, S_i, Y_i, X_i)$  for  $i = 1, \dots, N$ , be independent and identically distributed (i.i.d.) copies of  $O = (Z, R, S, Y, X)$ , where if  $R_i = 0$ ,  $S_i$  and  $Y_i$  are not included in the observed data vector.

A causal estimate of the treatment effect on severity of prostate cancer among patients who would have developed biopsy-detectable prostate cancer regardless of treatment assignment can be expressed as a risk ratio, odds ratio, or absolute risk difference; we use the latter, referred to here as the average causal effect:

$$ACE \equiv E(Y(1) - Y(0) | S(0) = S(1) = 1).$$

Within each treatment arm the group of subjects who developed cancer is a mixture of subjects who would have always had cancer and those who would not have had cancer had they received the other treatment; see Table 2. It is important to note therefore that  $ACE$  is not necessarily equal to the difference of observable conditional expectations,  $E(Y | S = 1, R = 1, Z = z)$ , for  $z = 0, 1$ .

We will make the independence assumption

$$(R(0), R(1), S(0), S(1), Y(0), Y(1), X(0), X(1)) \perp\!\!\!\perp Z, \quad (2)$$

which is ensured by randomization. Under (1) and (2),  $E(Y(z)|R(z) = 1, S(z) = 1) = E(Y | R = 1, S = 1, Z = z)$ . Because we do not know who would have developed cancer regardless of treatment assignment, *ACE* is not identifiable under (1) and (2) alone, and requires additional assumptions for estimation.

## 2.2 Assumptions to Identify the Average Causal Effect

In a first analysis to identify *ACE* we make the following additional assumptions:

$$Y(z), S(z) \perp\!\!\!\perp R(z), \quad (3)$$

$$S(0) \geq S(1), \quad (4)$$

$$Y(0) \perp\!\!\!\perp S(1) | S(0) = 1, \quad (5)$$

where for random variables  $A, B$ , and  $C$ ,  $A \perp\!\!\!\perp B | C$  indicates conditional independence of  $A$  and  $B$  given  $C$ . Assumption (3) states that obtaining a biopsy is independent of cancer status and severity, and is equivalent to the assumption that cancer status is MCAR. Under this assumption,  $E(Y(z) | S(z) = 1, R(z) = 1) = E(Y(z) | S(z) = 1)$ , that within treatment arm the average risk of high-grade cancer among cancers for men whose cancer status was known equals the average risk of high grade cancer among cancers for men whether or not their cancer status was known.

Assumption (4) is often referred to as monotonicity and implies that everyone who developed biopsy-detectable prostate cancer in the finasteride arm also would have developed biopsy-detectable cancer if randomized to placebo. Under this assumption, the probability that a participant who developed cancer under placebo would have developed cancer had they instead been randomized to receive finasteride,  $P(S(1) = 1 | S(0) = 1)$ , is equal to  $P(S(1) = 1) / P(S(0) = 1)$ , which under (1) and (2) can be estimated as the relative risk of cancer. While assumption (4) is not consistently testable, there was no evidence suggesting its violation since finasteride appeared to have a beneficial effect of partially preventing prostate cancer across all covariate-defined subgroups (Thompson et al. 2003).

Assumption (5) states that among subjects who developed prostate cancer in the placebo arm, their cancer status had they been randomized to finasteride is independent of the observed severity of their cancer. In other words, the placebo arm distribution of the severity of prostate cancer is the same in the always diseased principal stratum ( $S(0) = S(1) = 1$ ; stratum CC in Table 2) and the protected principal stratum ( $S(0) = 1, S(1) = 0$ ; stratum CN).

Assumptions (1)–(5) and the observed data identify *ACE*

$$\begin{aligned}
 ACE &= E(Y(1)|S(1)=1) - E(Y(0)|S(0)=1, S(1)=1) \\
 &= E(Y(1)|S(1)=1) - E(Y(0)|S(0)=1) \\
 &= E(Y(1)|R(1)=1, S(1)=1) - E(Y(0)|R(1)=1, S(0)=1) \\
 &= E(Y|R=1, S=1, Z=1) - E(Y|R=1, S=1, Z=0)
 \end{aligned}$$

where the first line is by (4), the second by (5), the third by (3), and the final by (1) and (2). The *ACE* for the PCPT data is therefore estimated as  $299/821 - 264/1194 = 0.14$  (95% Wald confidence interval of 0.10, 0.18), indicating that finasteride caused a statistically significant 14% absolute risk increase in high-grade prostate cancer compared to the placebo arm.

Assumptions (3)–(5) are not consistently testable from empirical data and are refutable in the PCPT. These assumptions are not the only assumptions that identify *ACE* as equal to the difference of observable conditional expectations. We consider (3)–(5) because they provide a reasonable and interpretable platform for facilitating a sensitivity analysis in the context of the PCPT.

### 3. SENSITIVITY ANALYSES

#### 3.1 Relaxing Assumption (5)

Assumption (5) states that the grade of prostate cancer among cases on placebo is unrelated to whether or not they would have developed cancer had they been on finasteride. This assumption may be implausible since finasteride may well be less effective against more aggressive prostate cancer.

Gilbert, Bosch, and Hudgens (2003) (GBH) proposed a flexible approach for relaxing (5) by assuming the model

$$\text{logit } P(S(1)=1|S(0)=1, Y(0)=y) = \alpha_0 + \beta_0 y, \quad (6)$$

where  $\text{logit } a = \log(a/(1-a))$ ,  $\alpha_0$  is an unknown parameter, and  $\beta_0$  is fixed and known. For men who developed cancer on placebo, the logistic model (6) uses a sensitivity parameter  $\beta_0$  to link observed cancer grade to the probability of developing cancer if on finasteride. The  $\beta_0$  is interpreted as the difference in log-odds (so  $\exp(\beta_0)$  is the odds ratio) of prostate cancer on the finasteride arm for high- versus low-grade prostate cancer on the placebo arm. The  $\beta_0$  is not identifiable from the observed data; it is a fixed sensitivity parameter and varied as part of a sensitivity analysis. Setting  $\beta_0 = 0$  corresponds to the conditional independence assumption (5). Setting  $\beta_0 = \pm\infty$  corresponds to the bounds for *ACE* given by Hudgens, Hoering, and Self (2003) (HHS), and implies that those diagnosed with cancer in placebo with the most (or least) severe cancer are those who would have been diagnosed with cancer if randomized to finasteride.

Under (1)–(4) and (6), GBH proposed estimating  $\alpha_0$  using the fact that  $P(S(1)=1|S(0)=1) = \sum_{y=0,1} P(S(1)=1|S(0)=1, Y(0)=y)P(Y(0)=y|S(0)=1)$ , and by recognizing that  $P(S(1)=1|S(0)=1)$  can be estimated as the observed relative risk of prostate cancer (discussed in section 2.2) and that  $P(Y(0)=y|S(0)=1)$  can be estimated as the observed proportion of placebo cancer cases that are high- /low-grade. Once  $\alpha_0$  has been estimated, the probability of high-grade cancer for placebos who would have developed cancer under either treatment is estimated by plugging in estimates to the expectation of the biased sample model:



$$E(Y(0)|S(0)=S(1)=1) = \frac{\sum_{y=\{0,1\}} y P(S(1)=1|S(0)=1, Y(0)=y) P(Y(0)=y|S(0)=1)}{P(S(1)=1|S(0)=1)}.$$

Given that under (1)–(4),  $E(Y(1)|S(0)=S(1)=1) = E(Y|R=1, S=1, Z=1)$  (discussed in section 2.2), estimation of *ACE* follows, and the variance can be estimated via the bootstrap or as described in the Appendix.

We elicited plausible ranges for  $\beta_0$  from two subject matter experts from independent institutions, one a clinician who is particularly enthusiastic concerning finasteride treatment and the other an epidemiologist who has been more pessimistic. We prompted these experts with the question, “Given two men assigned placebo who got cancer during the course of the trial: who do you believe would be more likely to have gotten cancer if, contrary to fact, they were assigned finasteride? \_\_\_ the person with the higher Gleason score, \_\_\_ the person with the lower Gleason score, or \_\_\_ the two are equally likely.” We then elicited odds ratios and ranges. Both experts felt that the man with high Gleason on placebo would more likely have developed prostate cancer on finasteride, and provided ranges for the odds ratio  $\exp(\beta_0)$  of (1.05, 1.35) and (2.00, 4.00), respectively. Note that neither interval contains 1, implying that neither expert believed assumption (5) scientifically plausible.

Figure 2 shows estimation of *ACE* under (1)–(4), and (6) repeated for  $\beta_0$  in the broad interval  $[-5, 5]$ , i.e. odds ratio  $\exp(\beta_0) \in [.007, 148]$ , with 95% Wald confidence intervals constructed using the asymptotic expression for the variance of the estimated *ACE*. (Confidence intervals were similar for percentile bootstrap intervals based on 1000 replications.) Figure 2 also includes estimates and percentile bootstrap confidence intervals for  $\beta_0 = \pm\infty$ . For this analysis the elicited ranges did not come into play since for all  $\beta_0$ , including  $\pm\infty$ , the null hypothesis that *ACE* = 0 was rejected at the 0.05 level. Therefore under (1)–(4) and (6), no matter what the hypothesized relationship between high-grade prostate cancer on placebo and the risk of prostate cancer on finasteride, among people who would have had prostate cancer detected on biopsy on either arm of the study, those on finasteride had a statistically significant higher risk of developing high-grade prostate cancer.

### 3.2 Relaxing Assumptions (4) and (5)

The assumption of monotonicity (4), which states that everyone with cancer in finasteride would have gotten cancer if randomized to placebo, is strong and may not be plausible (Dawid, 2000), even though finasteride appeared to reduce the risk of prostate cancer across all trial subgroups. In a Ph.D. dissertation with Andrea Rotnitzky, Jemai (2005) proposed sensitivity analysis methods that relaxed (4) in addition to (5). Following their arguments, it can be shown that in addition to (1), (2), and (3), *ACE* is identified under (6) and the following two assumptions:

$$P(S(0)=1|S(1)=1) = \phi, \tag{7}$$

$$\text{logit } P(S(0)=1|S(1)=1, Y(1)=y) = \alpha_1 + \beta_1 y, \tag{8}$$

where  $\alpha_0$  and  $\alpha_1$  are unknown parameters to be estimated and  $\phi$ ,  $\beta_0$ , and  $\beta_1$  are fixed sensitivity parameters to be varied as part of the sensitivity analysis. By relaxing monotonicity, we no longer assume that all men with detectable cancer on finasteride would have developed detectable cancer on placebo. Assumption (7) specifies the probability of getting detectable

cancer in placebo given detectable cancer in finasteride and assumption (8), which is analogous to (6), links this probability to cancer grade. The sensitivity parameter  $\beta_1$  has a similar definition to  $\beta_0$  described in Section 3.1. In the PCPT, the sensitivity parameter  $\phi$  may lie anywhere between 0 and 1, with  $\phi = 1$  corresponding to monotonicity,  $\phi = P(S(0) = 1)$  corresponding to independence between  $S(0)$  and  $S(1)$ , and  $\phi = 0$  implying that no one with cancer on finasteride would have developed cancer if on placebo. For fixed  $\phi$ ,  $\beta_0$  and  $\beta_1$ , estimating equations analogous to those of Jemai's dissertation used to estimate  $ACE$  are shown in the Appendix.

For performing a sensitivity analysis we first note that bounds on  $ACE$  can be constructed by not restricting any of the sensitivity parameters. When this is done the bounds on  $ACE$  are  $-1$  and  $1$ , the minimum and maximum possible values of  $ACE$  without even looking at the data, so that these bounds are uninformative. Note that the method proposed by Zhang and Rubin (2003) would obtain identical bounds. Therefore to extract any useful information from this analysis it is necessary to establish plausible ranges of the sensitivity parameters as done in Section 3.1.

In addition to eliciting a plausible range for  $\beta_0$  (discussed in the previous section), we also elicited plausible ranges for  $\beta_1$  and  $\phi$  from our two subject matter experts. Our pessimist chose the ranges  $\phi \in [0.8, 0.95]$ ,  $\exp(\beta_0) \in [2, 4]$ , and  $\exp(\beta_1) \in [0.25, 0.50]$ . These ranges reflect a belief that monotonicity is slightly violated (between 5 and 20% of those with detectable cancer in the finasteride arm would not have gotten detectable cancer if randomized to placebo), that those with more severe forms of cancer in the placebo arm are more likely those who would have gotten detectable cancer if randomized to the finasteride arm, and that those with less severe forms of cancer in the finasteride arm are those who are more likely to have gotten detectable cancer if randomized to the placebo arm. Our optimist chose the ranges  $\phi \in [1.0, 1.0]$ ,  $\exp(\beta_0) \in [1.05, 1.35]$ , and  $\exp(\beta_1) \in [1.05, 1.35]$ . Notice from his range for  $\phi$  that this expert believed the monotonicity assumption. Therefore, his range for  $\beta_1$  is irrelevant, because in (8),  $P(S(0) = 1 | S(1) = 1, Y(1) = y) = 1$  irrespective of the value of  $y$ , and therefore the analysis of Section 3.1 (Figure 2) is the appropriate sensitivity analysis of  $ACE$  corresponding to this expert's opinions.

Figure 3 shows a sensitivity analysis of  $ACE$ , varying  $\phi$ ,  $\beta_0$ , and  $\beta_1$ . It seems unlikely that developing cancer on the finasteride arm is independent of developing cancer on the placebo arm ( $\phi = 0.24$ ). Even more unlikely is that  $S(0)$  and  $S(1)$  are negatively correlated, so plots assuming  $\phi < 0.24$  are not shown. For display purposes the range for  $\beta_0$  (and  $\beta_1$ ) was condensed from  $[-5, 5]$  in Figure 2 to  $[-2.5, 2.5]$  in Figure 3, corresponding to odds ratios between 0.08 and 12.2. Contours represent the estimated  $ACE$  at a given  $\phi$ ,  $\beta_0$ , and  $\beta_1$ . Shaded regions correspond to those sensitivity parameter values where the Wald-based 95% confidence interval for  $ACE$  does not contain 0. Therefore, estimates in the dark-shaded regions imply that among those who would have gotten cancer regardless of treatment assignment finasteride caused high-grade cancer, the light-shaded regions imply that finasteride lowered cancer grade, and estimates in the unshaded region fail to reject  $H_0 : ACE = 0$  at the 0.05-level. Notice in Figure 3 that at  $\phi = 0.99$  (a minor violation of monotonicity),  $H_0 : ACE = 0$  is rejected for all  $\beta_0$  and  $\beta_1$  between  $[-2.5, 2.5]$ , consistent with the analysis performed under the assumption of monotonicity (Figure 2). As  $\phi$  moves farther from 1, the range of estimates for  $ACE$  increases (bottom two graphs of Figure 3). Our experts' ranges for  $\phi$ ,  $\beta_0$ , and  $\beta_1$  are included as rectangles in Figure 3, with  $\times$  corresponding to their selections of the most likely values of the sensitivity parameters. (For illustrative purposes we have included our optimist's range for  $\beta_0$  and  $\beta_1$ , at  $\phi = 0.95$  and  $\phi = 0.99$ , although Figure 2, not Figure 3, is the appropriate sensitivity analysis corresponding to this expert's opinions.) The hypothesis  $H_0 : ACE = 0$  is rejected for most values of  $\phi$ ,  $\beta_0$ , and  $\beta_1$  within the experts' ranges; evidence that finasteride is causing higher grades of cancer. The exception is when  $\phi$  is at the far end of our pessimist's plausible range, e.g. if  $\phi = 0.8$ ,  $\exp(\beta_0) = 4$ , and  $\exp(\beta_1) = 0.25$  there is insufficient evidence to reject  $H_0$ .



### 3.3 Relaxing Assumptions (3), (4), and (5)

Assumption (3), that receipt of a biopsy is independent of cancer status and severity is refutable in the PCPT, particularly because men were referred to biopsy based on their PSA and DRE. Verification bias in the PCPT could also have differed by treatment arm since sensitivity of PSA for prostate cancer detection on biopsy was higher on the finasteride arm. To relax (3), we assume

$$Y(z), S(z) \perp\!\!\!\perp R(z) | X(z), \quad (9)$$

that conditional on covariates (baseline and post-baseline), receipt of biopsy is independent of cancer status and severity. In the PCPT, assumption (9) is much more reasonable than (3), as the study rigorously measured covariates related to both receipt of a biopsy and to prostate cancer outcome: PSA and DRE at each yearly visit, family history of prostate cancer, age, race, and prior negative biopsies.

Our approach is to perform a sensitivity analysis similar to Section 3.2, only assuming (9) instead of (3). Under (1), (2), (6)–(9), *ACE* can be estimated by augmenting the estimating equations by weights corresponding to the inverse probability of having a biopsy, in a manner similar to that described by Robins, Rotnitzky, and Zhao (1995). Specifically, we use all participants to estimate the probability of having a biopsy based on covariates and treatment assignment,  $P(R = 1 | X = x, Z = z)$ ; the estimating equations for an individual who had a biopsy are then weighted by the inverse of their estimated probability of getting a biopsy.

In the PCPT many participants had biopsies during the course of the trial. Consistent with the original analysis, we have defined  $S = 1$  as a positive biopsy at any time during the trial,  $S = 0$  as a negative biopsy 7 years after randomization (as well as all prior biopsies), and  $R = 0$  if no biopsy was taken 7 years after randomization and all prior biopsies (if any) during the course of the study were negative. Therefore, cancer status within 7 years of randomization was unknown ( $R = 0$ ) for two reasons: 1) dropout before 7 years of follow-up without a cancer diagnosis, or 2) staying in the study for 7 years with no positive interim biopsy and deciding not to have an end-of-study biopsy. One can therefore think of  $P(R = 1 | X = x, Z = z) = P(A = B = 1 | X = x, Z = z) = P(B = 1 | A = 1, X = x, Z = z)P(A = 1 | X = x, Z = z)$ , where  $A$  is the indicator of staying in the study until a cancer status ( $S$ ) ascertaining biopsy and  $B$  is the indicator of having a biopsy. We modeled and estimated both probabilities separately, multiplying the two together to estimate  $P(R = 1 | X = x, Z = z)$ . The probability of not dropping out of the study before diagnosis of  $S$  was modeled using logistic regression with  $Z$  and covariates baseline PSA, age, race, and family history of prostate cancer. Given that  $A = 1$ , the probability of choosing to have a biopsy was modeled using logistic regression with  $Z$ , baseline PSA, age, race, family history of prostate cancer, abnormal DRE at last visit and its interaction with treatment, biopsy recommendation based on high PSA at last visit and its interaction with treatment, and prior negative biopsy during the course of the trial. Estimation details are given in the Appendix.

Results under (9) shown in Figure 4 are very similar to the results under (3) shown in Figure 3. The estimated *ACE* is slightly lower under (9) than under (3) for the same sensitivity parameter values, but this difference is not substantial. That results under (9) are similar to results under (3) is consistent with analyses of the receiver-operating characteristic (ROC) curves for both arms of the study, where Thompson and colleagues found no difference in the area under the ROC for analyses accounting for and not accounting for potential verification bias by incorporating biopsy-predicting covariates (Thompson et al. 2005, 2006).

## 4. DIFFERENTIAL BIOPSY GRADING

We re-performed the sensitivity analyses of Section 3.3 (assuming (1), (2), (6)–(9)), accounting for differential biopsy high-grade detection by using the prostatectomy results from the 531 participants in Table 4. Cancer grade was reported as missing for those diagnosed with cancer but without a prostatectomy. To account for potential bias due to our excluding those without a prostatectomy, we again employed inverse probability weights. Among those diagnosed with prostate cancer, we first estimated the probability of having a prostatectomy based on treatment, original cancer severity, and covariates using a linear logistic model. Next, the components of our estimating equations which incorporated cancer severity were multiplied by an indicator of getting a prostatectomy and the subject-specific inverse probability of prostatectomy estimated using the logistic model. Variance estimates were altered to account for this weighting. Details are in the Appendix.

Figure 5 is the resulting sensitivity analysis of  $ACE$ , using prostatectomy-defined cancer grades. Notice now that over the entire range of sensitivity parameters chosen by our experts, there is insufficient evidence to reject  $H_0 : ACE = 0$ . The inability to reject the null is due to two factors: First, for a given  $\beta_0$ ,  $\beta_1$ , and  $\phi$ , the estimated  $ACE$  decreased when using prostatectomy measures of cancer severity (Figure 5) compared to the original measurement (Figure 4). Second, the variance of the estimated  $ACE$  increased when using prostatectomy measures. This loss of power is not surprising, as prostatectomies were only obtained in about a quarter of those who were diagnosed with prostate cancer.

It should also be noted that results using inverse probability weights can be highly variable when there are extreme weights. The weights in this prostatectomy analysis were similar between treatment arms, but ranged from 1.7 to 49.2. There were two individuals with extreme weights ( $> 46$ ), one in finasteride with  $Y = 1$  and one in placebo with  $Y = 0$ . If these two individuals were removed the maximum weight was much smaller, 25.9, and for given sensitivity parameters, the estimated  $ACE$  and its standard error decreased (e.g., at  $\phi = 0.9$ ,  $exp(\beta_0) = 3$  and  $exp(\beta_1) = 1/3$ ,  $\widehat{ACE} = -0.068$  with standard error = 0.053 compared to  $\widehat{ACE} = -0.031$  with standard error 0.060 in the analysis of Figure 5). In general, conclusions were similar (data not shown).

## 5. DISCUSSION

So does finasteride affect the severity of prostate cancer? It depends on the assumptions one is willing to make. If one does not make any assumptions,  $ACE$  can take any value between  $-1$  and  $1$ , implying that one can draw any conclusion one would like from this data. Ignoring the prostatectomy results, over most of the sensitivity parameter ranges chosen by our subject matter experts, estimates and 95% confidence intervals for  $ACE$  were greater than 0, implying that finasteride does increase the severity of cancer (Figure 2–Figure 4). However, using the more recent prostatectomy measures of disease severity to account for potential bias due to differential biopsy grading, over all sensitivity parameter ranges chosen by our experts there was insufficient evidence to reject  $H_0 : ACE = 0$  (Figure 5). These latter estimates were less precise, as they were based on a smaller proportion of trial participants. More efficient methods of accounting for misclassification of  $Y$  warrant further study.

Our sensitivity analyses do not yield a single answer, which some might find unattractive. In contrast, we could have chosen a range of plausible sensitivity parameters, put a distribution on this range, and then integrated  $ACE$  over this distribution. Or more formally, we could have performed a fully Bayesian analysis, putting a prior on our sensitivity parameters and estimating the posterior distribution of  $ACE$  (e.g., Scharfstein et al. 2003). Such approaches are reasonable and may yield simpler answers. However, as the estimated  $ACE$  is highly

dependent on the sensitivity parameters and hence the choice of prior, we prefer to show results under a wide range of sensitivity parameters, letting the reader draw conclusions based on his/her personal beliefs.

Of course, in order to make sense of a sensitivity analysis of this type, interpretation of the sensitivity parameters is key. Although choosing a range for counterfactual sensitivity parameters can be challenging, we believe that our subject matter experts understood them. We specifically chose an expert who we thought would be an optimist and another expert who we thought would be a pessimist. Differences between the chosen sensitivity parameter ranges probably reflect differences of opinion rather than a poor understanding of the parameters. A discussion of the challenges of eliciting and interpreting similar sensitivity parameters, as well as a survey similar to the one we used to elicit our ranges is found elsewhere (Shepherd, Gilbert, and Mehrotra 2007). We recognize that other subject matter experts may have opinions very different than those elicited from our two experts. A more thorough picture of expert opinion about the sensitivity parameter would require elicitation from more experts. Although one can imagine experiments that may favor choosing certain ranges for the sensitivity parameters, it should be recognized that no experiment can truly estimate these parameters as they are indeed counterfactual.

In Section 3.3, we relaxed assumption (3) by assuming (9), that having a biopsy was independent of cancer status or severity conditional on covariates. Instead, we could have performed sensitivity analyses following the general approach of Rotnitzky et al. (1998) and assumed the following model for missing biopsy:

$$\text{logit } P(R(z)=1|S(z)=s, Y(z)=y)=\eta_0+\eta_1z+\tau_0s+\tau_1zs+\tau_2sy+\tau_3zsy, \quad (10)$$

for  $z = 0, 1, s = 0, 1,$  and  $y = 0, 1,$  where  $\eta = (\eta_0, \eta_1)$  are unknown scalars and  $\tau_0, \tau_1, \tau_2,$  and  $\tau_3$  are sensitivity parameters. However, this approach introduces 4 new sensitivity parameters, bringing us to a total of 7 sensitivity parameters – too many to be of practical use – and we believe the PCPT gathered the right covariates to make (9) reasonable.

There are a few additional issues we did not address in our sensitivity analyses. The first is that biopsy does not perfectly detect prostate cancer and that detection may vary by treatment. Throughout this paper we have defined  $S$  as *biopsy-detectable* prostate cancer. Therefore, interpretation of  $ACE$  is limited to the effect of finasteride on severity of cancer among those with *biopsy-detectable* prostate cancer under either treatment. While it is reasonable to assume that a subject with cancer detected on biopsy truly has cancer, a negative biopsy does not necessarily mean that there is no cancer present. It is also likely that the negative predictive value (NPV) of biopsy for prostate cancer on finasteride is larger than the NPV on placebo because of finasteride's tendency to shrink the prostate. Hence, more cancers were likely missed on the placebo arm than on the finasteride arm. The implications of this potential differential cancer detection on the estimated  $ACE$  depend on the grade of those possibly undetected cancers. As we do not have this information nor estimates of the NPV on either treatment arm, this potential misclassification would have to be addressed by additional sensitivity analyses. Second, we have ignored compliance, so our conclusions can only be interpreted as the causal effect of randomization to finasteride, not actually taking finasteride. Baker (2000) proposed methods which address noncompliance in the context of the PCPT.

Finally, it should be noted that while prostate cancer grade is important, of greater clinical importance is whether finasteride decreased prostate cancer mortality. Unfortunately, the PCPT cannot answer this question because death by prostate cancer is rare; an answer would require longer follow-up and/or more participants. Perhaps the most clinically important

question the PCPT can answer is “Does finasteride reduce the risk of severe prostate cancer?” We have addressed a different question: “What is the effect of finasteride on cancer severity among those who would be diagnosed with cancer regardless of treatment?” Our question addresses the controversy of the PCPT’s conflicting results and finasteride’s causal mechanisms. However, our question is not as important from a clinical or public health perspective because it is not known which men will be diagnosed with cancer irrespective of treatment.

Although we have focussed on the PCPT, these methods are more generally applicable. One example is HIV vaccine trials where there is interest in estimating the effect of vaccination on post-infection outcomes among those who would have been infected regardless of treatment assignment (HHS, GBH). Proposed methods in this context have assumed monotonicity and have ignored missing data (missing infection status and/or missing post-infection outcome if infected). The methods presented here relax monotonicity, account for missing data, and can be applied without alteration when the outcome is continuous. Another possible application of these methods is for examining the causal effect of treatment on an outcome that only exists in survivors (e.g., Hayden, Pauler, and Schoenfeld 2005; Egleston et al. 2007).

In conclusion, our sensitivity analyses offer new insights about potential explanations of the increased number of high-grade cancers on the finasteride arm of the PCPT. Although not completely exhaustive, we believe they account for most of the potential biases that could artificially induce the conflicting results of an increased absolute number of high-grade prostate cancer cases on the finasteride arm in the face of a 25% reduction in biopsy-detectable prostate cancer. This finding appears not to be due to differential biopsy verification but could be due to the improved sensitivity of biopsy for detecting high grade disease in finasteride compared to placebo, which when accounted for removes the statistical significance of the average causal effect of increased high-grade prostate cancer by finasteride.

## Acknowledgments

We would like to thank Catherine Tangen, Phyllis Goodman, Ian Thompson, Alan Kristal, and William Dupont for their help with this manuscript. This article was supported in part by Public Health Service grant CA37429 from the National Cancer Institute.

## REFERENCES

- Baker SG. Analyzing a randomized cancer prevention trial with a missing binary outcome, an auxiliary variable, and all-or-none compliance. *Journal of the American Statistical Association* 2000;95:43–50.
- Cox, DR. *Planning of Experiments*. New York: Wiley; 1958.
- Egleston BL, Scharfstein DO, Freeman EE, West SK. Causal inference for non-mortality outcomes in the presence of death. *Biostatistics* 2007;8:526–545. [PubMed: 16980696]
- Frangakis CE, Rubin DB. Principal stratification in causal inference. *Biometrics* 2002;58:21–29. [PubMed: 11890317]
- Gilbert PB, Bosch RJ, Hudgens MG. Sensitivity analysis for the assessment of causal vaccine effects on viral load in HIV vaccine trials. *Biometrics* 2003;59:531–541. [PubMed: 14601754]
- Goodman PJ, Tangen CM, Crowley JJ, Carlin SM, Ryan A, Coltman CA Jr, Ford LG, Thompson IM. Implementation of the Prostate Cancer Prevention Trial (PCPT). *Controlled Clinical Trials* 2004;25:203–222. [PubMed: 15020037]
- Hayden D, Pauler DK, Schoenfeld D. An estimator for treatment comparisons among survivors in randomized trials. *Biometrics* 2005;61:305–310. [PubMed: 15737107]
- Hudgens MG, Halloran ME. Causal vaccine effects on binary postinfection outcomes. *Journal of the American Statistical Association* 2006;101:51–64. [PubMed: 19096723]
- Hudgens MG, Hoering A, Self SG. On the analysis of viral load endpoints in HIV vaccine trials. *Statistics in Medicine* 2003;22:2281–2298. [PubMed: 12854093]

- Jemai, Y. unpublished Ph.D. dissertation under the supervision of A. Rotnitzky. Harvard School of Public Health, Department of Biostatistics; 2005. Semiparametric methods for inferring treatment effects on outcomes defined only if a post-randomization event occurs.
- Little, RJA.; Rubin, DB. Statistical Analysis with Missing Data. New York: Wiley; 2002.
- Lucia MS, Epstein JI, Goodman PJ, Darke AK, Reuter VE, Civantos F, La Rosa F, Kattan MW, Tangen CM, Lippman SM, Parnes HL, Coltman CA, Thompson IM. Finasteride and high-grade prostate cancer in the Prostate Cancer Prevention Trial. *Journal of the National Cancer Institute* 2007;99:1375–1383. [PubMed: 17848673]
- Newey, WK.; McFadden, D. Large sample estimation and hypothesis testing. In: Engle, RF.; McFadden, DL., editors. *Handbook of Econometrics*. Vol. Volume IV. Elsevier Science B.V.; 1994.
- Neyman J. On the application of probability theory to agricultural experiments: Essay on principles. Translated in *Statistical Science* 1923;5:465–480.
- Robins JM. An analytic method for randomized trials with informative censoring: Part I. *Lifetime Data Analysis* 1995;1:241–254. [PubMed: 9385104]
- Robins JM, Rotnitzky A, Zhao LP. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association* 1995;90:106–121.
- Rosenbaum PR. The consequences of adjustment for a concomitant variable that has been affected by the treatment. *Journal of the Royal Statistical Society, Series A* 1984;147:656–666.
- Rotnitzky A, Robins JM, Scharfstein DO. Semiparametric regression for repeated outcomes with nonignorable nonresponse. *Journal of the American Statistical Association* 1998;93:1321–1339.
- Rubin DB. Bayesian inference for causal effects: the role of randomization. *The Annals of Statistics* 1978;6:34–58.
- Rubin, DB. Causal inference without counterfactuals. In: Dawid, AP., editor. *Journal of the American Statistical Association*. Vol. 95. 2000. p. 435–437. Comment on
- Scardino PT. The prevention of prostate cancer – the dilemma continues. *The New England Journal of Medicine* 2003;349:295–297.
- Scharfstein DO, Daniels MJ, Robins JM. Incorporating prior beliefs about selection bias into the analysis of randomized trials with missing outcomes. *Biostatistics* 2003;4:495–512. [PubMed: 14557107]
- Shepherd BE, Gilbert PB, Mehrotra DV. Eliciting a counterfactual sensitivity parameter. *The American Statistician* 2007;61:56–63.
- Thompson IM, Goodman PJ, Tangen CM, Lucia MS, Miller GJ, Ford LG, Lieber MM, Cespedes RD, Atkins JN, Lippman SM, Carlin SM, Ryan A, Szczepanek CM, Crowley JJ, Coltman CA. The influence of finasteride on the development of prostate cancer. *The New England Journal of Medicine* 2003;349:213–222. [PubMed: 12867605]
- Thompson IM, Ankerst DP, Chi C, Lucia MS, Goodman P, Crowley JJ, Parnes HL, Coltman CA. The operating characteristics of prostate-specific antigen in a population with initial PSA of 3.0 ng/ml or lower. *Journal of the American Medical Association* 2005;294:66–70. [PubMed: 15998892]
- Thompson IM, Chi C, Ankerst DP, Goodman P, Tangen C, Lippman S, Lucia MS, Parnes HL, Coltman CA. Effect of finasteride on the sensitivity of PSA for detecting prostate cancer. *Journal of the National Cancer Institute* 2006;98:1128–1133. [PubMed: 16912265]
- Thompson IM, Tangen CM, Goodman PJ, Lucia MS, Parnes HL, Lippman SM, Coltman CA. Finasteride improves the sensitivity of digital rectal examination for prostate cancer detection. *Journal of Urology* 2007;177:1749–1752. [PubMed: 17437804]
- Zhang JL, Rubin DB. Estimation of causal effects via principal stratification when some outcomes are truncated by ‘death, ’. *Journal of Educational and Behavioral Statistics* 2003;28:353–368.

## APPENDIX: ESTIMATING EQUATIONS

### A.1 Estimating equations for Section 3.2

Let  $\mu_z \equiv E(Y(z)|S(0) = S(1) = 1)$  and  $p_z \equiv P(S(z) = 1)$  for  $z = 0, 1$ . Therefore  $ACE = \mu_1 - \mu_0$ . Define  $\theta \equiv (p_0, p_1, \alpha_0, \alpha_1, \mu_0, \mu_1)$ , where  $\alpha_0, \alpha_1$  are given by (6), (8). Our sensitivity parameters are  $\beta_0, \beta_1$ , and  $\phi$ . Let  $N$  denote the total number of subjects with an end-of-study biopsy (i.e., the analysis is only performed on those with  $R = 1$ ).



To estimate  $\theta$  under (1), (2), (3), (6), (7), and (8), Jemai (2005) proposed an estimating equation of the form,  $\sum_{i=1}^N U_i(\theta)=0$ , Where

$$U_i(\theta)=\begin{cases} (1 - Z_i) (p_0 - S_i) \\ Z_i(p_1 - S_i) \\ (1 - Z_i) S_i \left\{ \frac{1}{1+\exp(-\alpha_0-\beta_0 Y_i)} - \frac{\phi p_1}{p_0} \right\} \\ Z_i S_i \left\{ \frac{1}{1+\exp(-\alpha_1-\beta_1 Y_i)} - \phi \right\} \\ (1 - Z_i) S_i \left\{ \mu_0 - Y_i \frac{1}{1+\exp(-\alpha_0-\beta_0 Y_i)} \frac{p_0}{\phi p_1} \right\} \\ Z_i S_i \left\{ \mu_1 - Y_i \frac{1}{1+\exp(-\alpha_1-\beta_1 Y_i)} \frac{1}{\phi} \right\} . \end{cases}$$

Jemai showed that the resulting estimate,  $\hat{\theta}$  is asymptotically normal with

$$\sqrt{N} (\hat{\theta} - \theta) \rightarrow^d N(0, C), \text{ where } C = \Gamma^{-1} \Omega \Gamma^{-1}, \Gamma = E \left[ \frac{\partial}{\partial \theta} U(\theta) \right], \text{ and } \Omega = E[U(\theta)U(\theta)'] . C \text{ is}$$

estimated in the usual manner by replacing expectations with  $N^{-1} \sum_{i=1}^N$  and plugging in  $\hat{\theta}$  for  $\theta$ . The variance of  $\widehat{ACE}$  is estimated as  $(\hat{C}_{55} + \hat{C}_{66} - 2 \hat{C}_{56})/N$ , where  $\hat{C}_{ij}$  corresponds to the  $i$ th row and  $j$ th column of the estimate of  $C$ . Estimates and confidence intervals for other smooth functions of  $\theta$ , such as the causal risk ratio  $\mu_1 / \mu_0$ , can be similarly constructed through an application of the continuous mapping theorem.

### Computation is straightforward

$p_0$  and  $p_1$  are first estimated and then plugged into the 3rd and 4th lines of  $U_i(\theta)$  to estimate  $\alpha_0$  and  $\alpha_1$  separately, using a one-dimensional optimizer (e.g., R's function **optimize**). Using the estimates of  $p_0, p_1, \alpha_0$ , and  $\alpha_1$ , the remaining parameters  $\mu_0$  and  $\mu_1$  are estimated. This results in the unique solution of the estimating equation.

Note that the approach of GBH employed in Section 3.1 is equivalent to solving  $\sum U_i(\theta) = 0$  with  $U_i(\theta)$  as defined above only setting  $\phi = (1+\exp(-\alpha_1 - \beta_1 Y_i))^{-1} = 1$ , and removing the fourth line of  $U_i(\theta)$  because  $\alpha_1$  is no longer estimated.

## A.2 Estimating equations for Section 3.3

Let  $N = 15991$  (i.e., we are including all participants, both  $R = 0$  and  $R = 1$ ). Let  $V_i(\theta, \hat{\eta}) = U_i(\theta)\lambda_i(\hat{\eta})^{-1} R_i$ , where  $\lambda_i(\eta) = P(R_i = 1|Z_i = z, X_i = x) = P(B_i = 1|A_i = 1, X_i = x, Z_i = z)P(A_i = 1|X_i = x, Z_i = z) = \lambda_{ai}(\eta_a)\lambda_{bi}(\eta_b)$  with  $\eta = (\eta_a, \eta_b)$ ,  $\hat{\eta}$  is the solution to the score equation,

$$N^{-1} \sum_{i=1}^N m_i(\eta) = 0, \text{ and } m_i(\eta) \text{ is the first derivative with respect to } \eta \text{ of the logarithm of } \lambda_{ai}(\eta_a)^{A_i} (1 - \lambda_{ai}(\eta_a))^{1-A_i} (\lambda_{bi}(\eta_b))^{B_i} (1 - \lambda_{bi}(\eta_b))^{1-B_i} A_i . \text{ We employ the models}$$

$$\lambda_{ai}(\eta_a) = (1 + \exp(-\eta'_a(1, Z_i, X_i)))^{-1} \text{ where } X_i = (\text{baseline PSA, age, age}^2, \text{white race indicator, African American indicator, indicator of family history of prostate cancer}) \text{ and}$$

$$\lambda_{bi}(\eta_b) = (1 + \exp(-\eta'_b(1, Z_i, X_i)))^{-1} \text{ where } X_i = (\text{baseline PSA, age, age}^2, \text{white race indicator, African American indicator, indicator of family history of prostate cancer, indicator of biopsy referral due to PSA, } Z \times (\text{indicator of biopsy referral due to PSA}), \text{ last DRE, } Z \times (\text{last DRE}), \text{ indicator of prior negative biopsy}). \text{ From Newey and McFadden (1994),}$$

$$\sqrt{N} (\hat{\theta} - \theta) \rightarrow^d N(0, C), \text{ where}$$

$$C = G_\theta^{-1} E[(V(\theta) + G_\eta \psi)(V(\theta) + G_\eta \psi)'] G_\theta^{-1}, G_\theta = E \left[ \frac{\partial}{\partial \theta} V_i(\theta, \eta) \right], G_\eta = E \left[ \frac{\partial}{\partial \eta} V_i(\theta, \eta) \right], \psi = -M^{-1} m(\eta),$$



and  $M = E \left[ \frac{\partial}{\partial \eta} m(\eta) \right]$ .  $C$  is estimated in the usual way, replacing expectations with  $N^{-1} \sum_{i=1}^N$  and plugging in  $\hat{\theta}$  and  $\hat{\eta}$  for  $\theta$  and  $\eta$ , respectively.

### A.3 Estimating equations for Section 4

Define  $Q_i$  as the indicator of subject  $i$  obtaining a prostatectomy. Re-define  $Y_i$  as the cancer high-grade indicator (0 or 1) based on the prostatectomy. Let  $Y_i^*$  be the previous cancer high-grade indicator based on biopsy. Model the probability of prostatectomy given prostate cancer with  $\lambda_{qi}(\eta_q) = (1 + \exp(-\eta_q'(1, Z_i, Y_i^*, X_i)))^{-1}$  where  $X_i =$  (baseline PSA, age, white race indicator, African American indicator, indicator of family history of prostate cancer, indicator of biopsy referral due to PSA, last DRE, indicator of prior negative biopsy). Let  $\hat{\eta}^q$  denote the estimate of  $\eta_q$ . Define

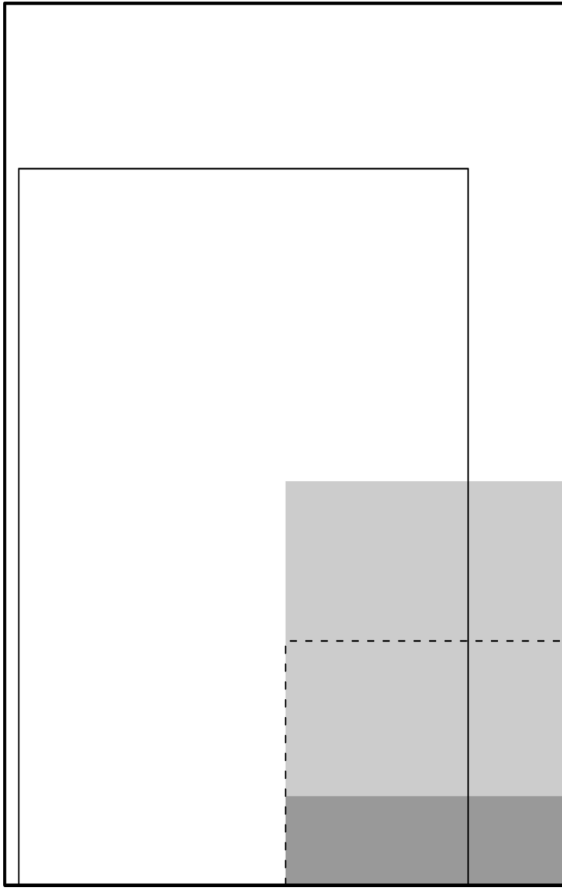
$$W_i(\theta, \hat{\eta}) = (V_{1i}(\theta, \hat{\eta}), V_{2i}(\theta, \hat{\eta}), V_{3i}(\theta, \hat{\eta}) \lambda_{qi}(\hat{\eta}_q)^{-1} Q_i, \dots, V_{6i}(\theta, \hat{\eta}) \lambda_{qi}(\hat{\eta}_q)^{-1} Q_i). \quad (11)$$

where  $(V_{1i}(\theta, \hat{\eta}), \dots, V_{6i}(\theta, \hat{\eta})) = V_i(\theta, \hat{\eta})$  as defined in A.2. Estimation is obtained by solving  $\sum_{i=1}^N W_i(\theta, \hat{\eta}, \hat{\eta}_q) = 0$ , where  $N = 15991$ . Notice that this estimating equation is the same as that given in A.2 except the last four components of the equation used to estimate  $(\alpha_0, \alpha_1, \mu_0, \mu_1)$  only use prostatectomy results, weighting by the inverse probability of having a prostatectomy.

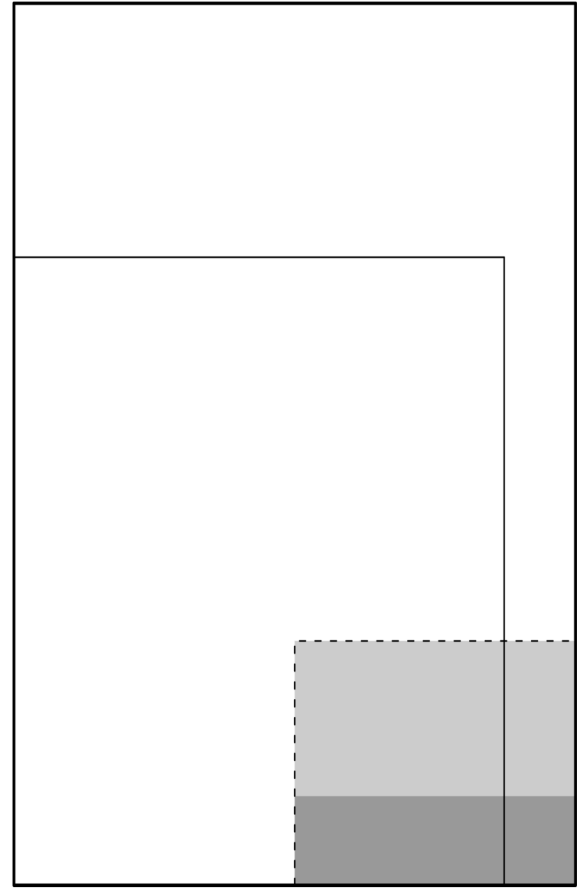
The variance is estimated in the same manner as described in A.2, except we now replace  $\eta$  of A.2 with  $\eta = (\eta_a, \eta_b, \eta_q)$ , and  $m_i(\eta)$  is the first derivative of the logarithm of

$\lambda_{ai}(\eta_a)^{A_i} (1 - \lambda_{ai}(\eta_a))^{1-A_i} (\lambda_{bi}(\eta_b))^{B_i} (1 - \lambda_{bi}(\eta_b))^{1-B_i} (\lambda_{qi}(\eta_q))^{Q_i} \lambda_{qi}(\eta_q)^{1-Q_i} \lambda_{qi}(\eta_q)^{A_i B_i Q_i}$ , and  $V(\cdot)$  is replaced with  $W(\cdot)$ .

## Placebo

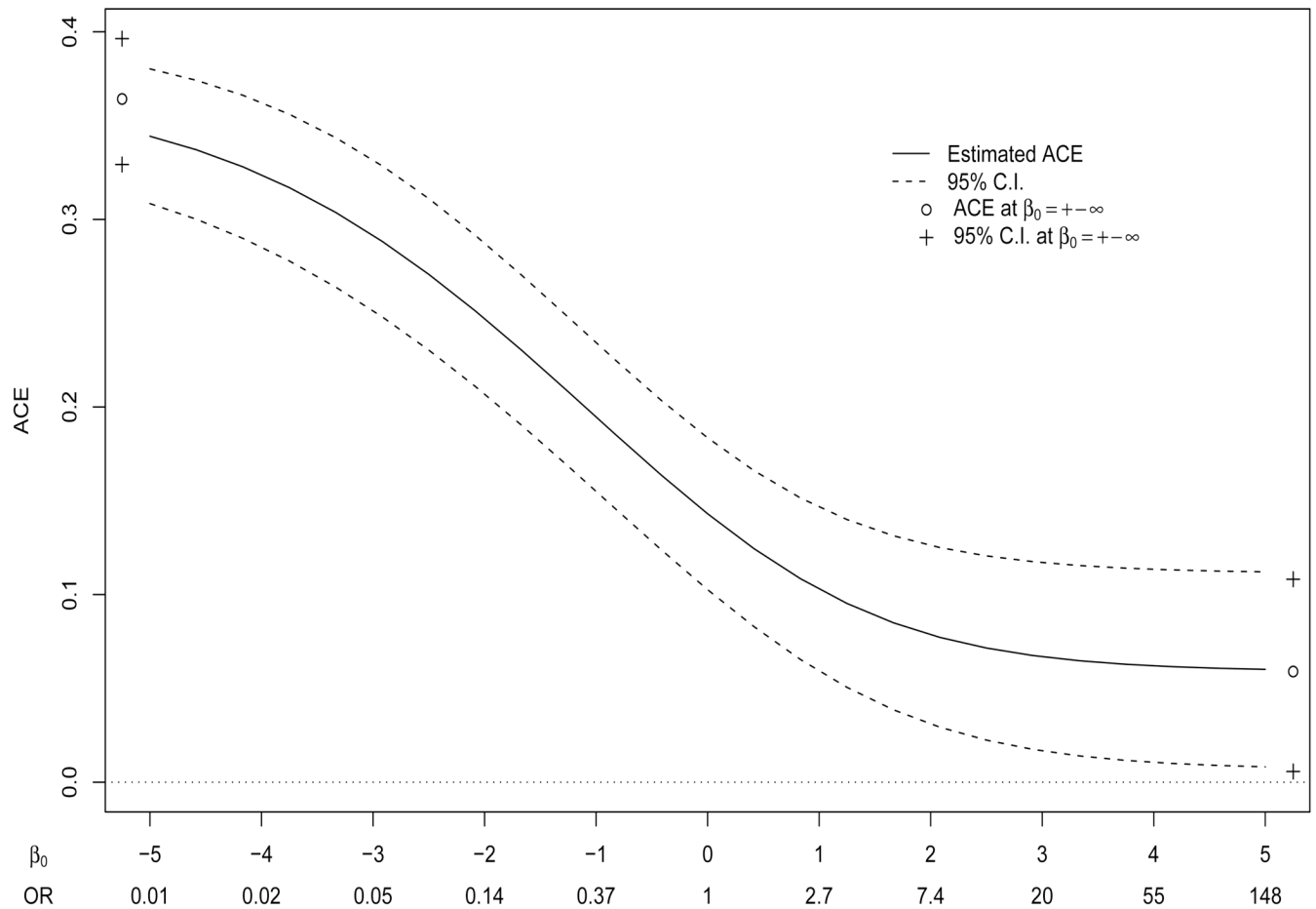


## Finasteride

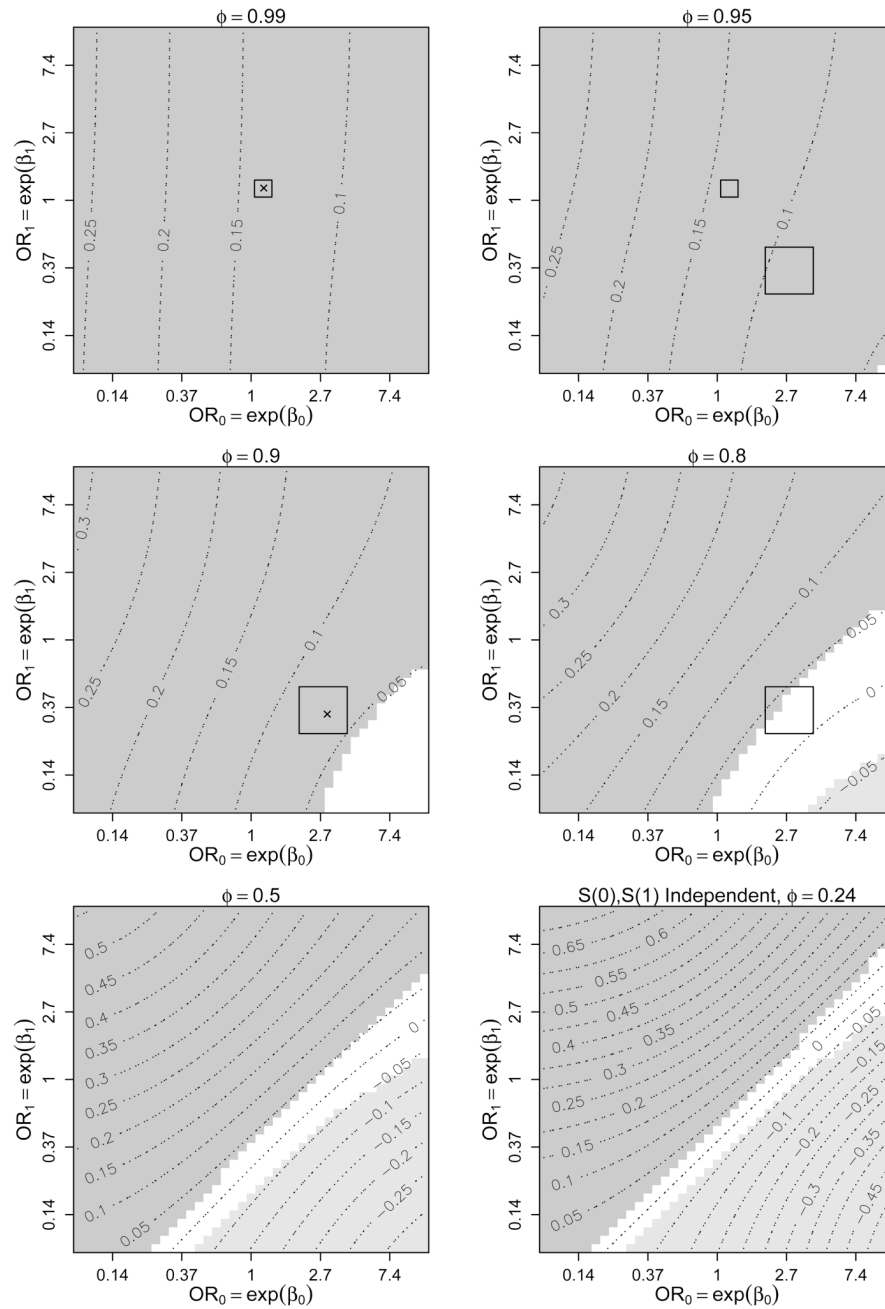


**Figure 1.**

Hypothetical example of the impact of post-randomization selection in the PCPT analysis. Because of randomization, men in the placebo and finasteride arms are comparable. The shaded regions are those men who developed prostate cancer, the darker shade corresponds to those who developed high-grade cancer. The solid inner boxes in both arms designate those men whose cancer status was determined by biopsy. Notice that the overall proportion of men who develop high-grade cancer is identical between treatment groups. However, the proportion of high-grade cancers among those diagnosed with cancer is much higher in the finasteride arm. The dashed box corresponds to those men who would have gotten cancer regardless of their treatment assignment; a comparison of cancer grade among men in this subpopulation allows one to assess the effect of finasteride on cancer severity.

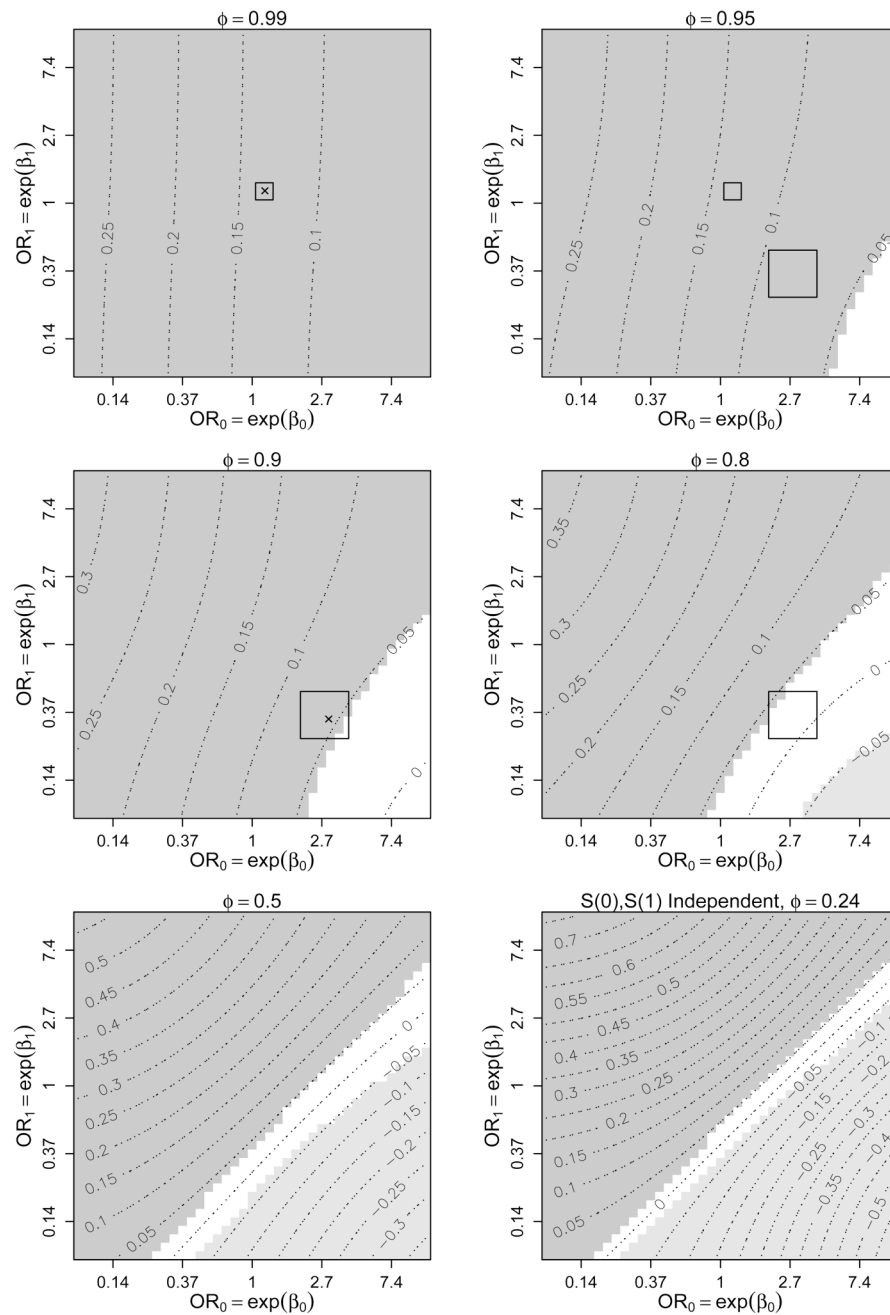


**Figure 2.** Sensitivity analysis of the average causal effect (ACE) of finasteride on severity of prostate cancer under (1), (2), and (3)–(6). The sensitivity parameter is given in terms of  $\beta_0$  ( $\log(\text{OR})$ ) and as odds ratios (OR).



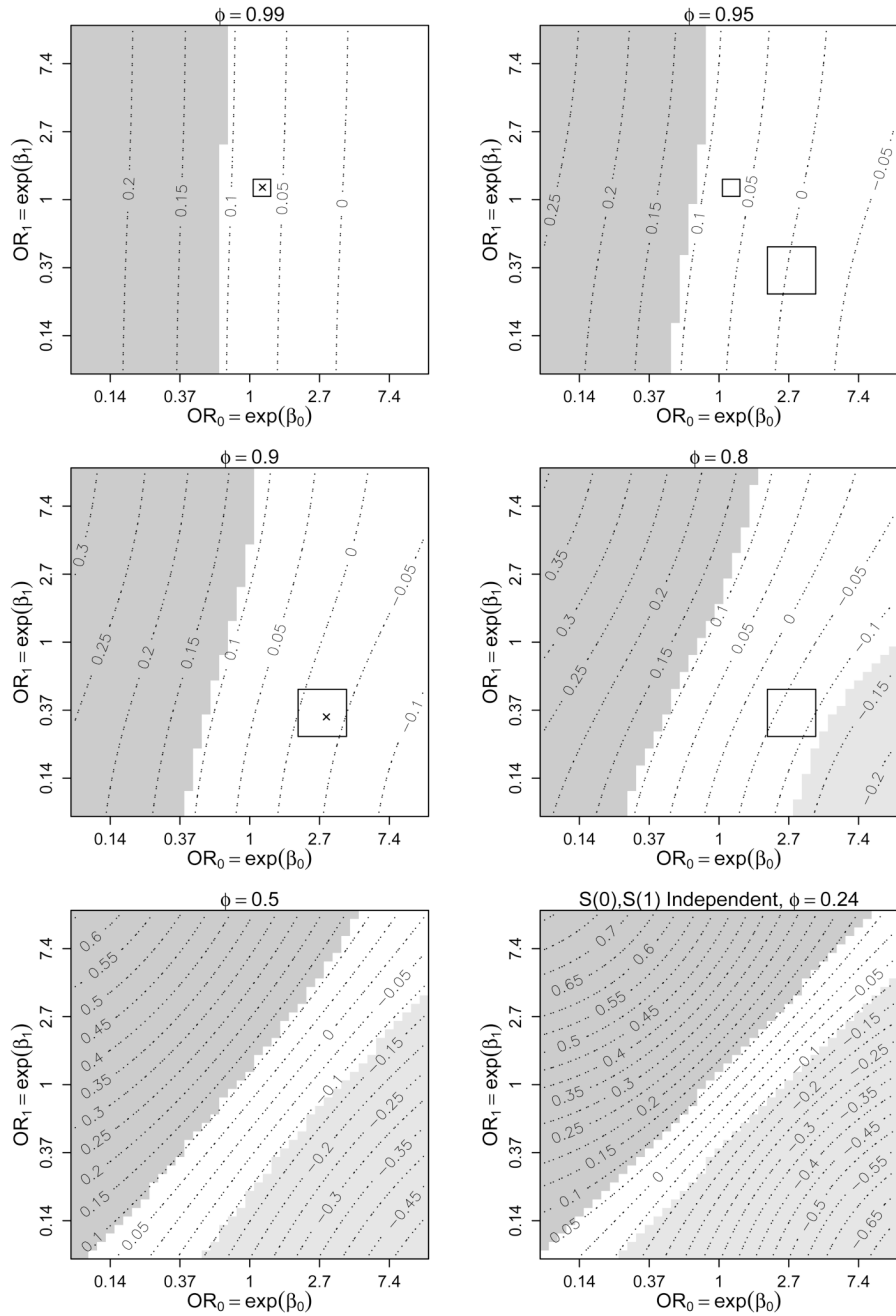
**Figure 3.** Contour plots of  $\widehat{ACE}$  for varying levels of the sensitivity parameters  $\beta_0$ ,  $\beta_1$ , and  $\phi$  under (1), (2), (3), and (6)–(8).  $\phi$  is the probability of prostate cancer if randomized to placebo given prostate cancer in the finasteride arm.  $OR_0$  ( $OR_1$ ) is the odds ratio for prostate cancer in finasteride (placebo) based on cancer grade, given cancer in placebo (finasteride). The shaded regions represent those choices of  $\beta_0$ ,  $\beta_1$ , and  $\phi$  where  $H_0 : ACE = 0$  is rejected at the 0.05-level. The darker (lighter) shades are where  $ACE > 0$  ( $ACE < 0$ ), implying a higher (lower) risk of high-grade prostate cancer on finasteride. The rectangles represent ranges of our sensitivity

parameters selected by two subject matter experts. The  $\times$  represent their most likely sensitivity parameter values.



**Figure 4.** Contour plots of  $\widehat{ACE}$  for varying levels of  $\beta_0$ ,  $\beta_1$ , and  $\phi$  under (1), (2), and (6)–(9). See Figure 3’s caption.





**Figure 5.** Contour plots of  $\widehat{ACE}$  using cancer grade based on prostatectomy for varying levels of  $\beta_0$ ,  $\beta_1$ , and  $\phi$  under (1), (2), and (6)–(9). See Figure 3’s caption.

**Table 1**

Numbers with biopsy, cancer, and high-grade cancer in the Prostate Cancer Prevention Trial.

	Placebo ( $Z = 0$ )	Finasteride ( $Z = 1$ )	Combined
Randomized	8025	7966	15991
Cancer Status Known by Biopsy ( $R = 1$ )	5217	4951	10168
Prostate Cancer ( $S = 1$ )	1194	821	2015
High Grade Prostate Cancer ( $Y = 1$ )	264	299	563

NOTE: These numbers are slightly different from those reported in Thompson et al. (2003) because of early study termination and because we only used participants with complete covariate information. We repeated the analyses reported in Sections 2.2, 3.1, and 3.2 using Thompson and colleagues' final dataset, and results were nearly identical.

**Table 2**

Principal Strata in Prostate Cancer Prevention Trial.

<b>Principal Stratum</b>	<b>Outcome on Placebo</b>	<b>Outcome on Finasteride</b>
NN	no cancer	no cancer
CN	cancer	no cancer
NC	no cancer	cancer
CC	cancer	cancer

**Table 3**

Characteristics of those with and without a study endpoint

	No Biopsy ( <i>R</i> = 0) ( <b>N=5823</b> )	Biopsy ( <i>R</i> = 1) ( <b>N=10168</b> )	P-value <sup>a</sup>
Finasteride Arm	3015 (51.8%)	4951 (48.7%)	0.0002
Baseline PSA	1.0 (0.6, 1.7) <sup>b</sup>	1.1 (0.7, 1.8)	< 0.0001
Baseline Age	63 (58, 67)	62 (58, 66)	0.0004
Race			< 0.0001
White	5310 (91.2%)	9471 (93.1%)	
African American	242 (4.2%)	346 (3.4%)	
Other	271 (4.7%)	351 (3.5%)	
Family History of Prostate Cancer	785 (13.5%)	1695 (16.7%)	< 0.0001
Referred to Biopsy because of PSA	69 (1.2%)	802 (7.9%)	< 0.0001
Referred to Biopsy because of DRE	83 (1.4%)	831 (8.2%)	< 0.0001
Prior Negative Biopsy	477 (8.2%)	1348 (13.3%)	< 0.0001

<sup>a</sup> Chi-square test for categorical variables, Wilcoxon rank sum test for continuous covariates (baseline PSA and age).

<sup>b</sup> Median (1st quartile, 3rd quartile) reported for continuous covariates.

**Table 4**  
Comparison of Cancer Grade based on Blinded Biopsy and Blinded Prostatectomy

	Biopsy:			
	Low Grade	High Grade	Total	
Placebo				
Prostatectomy				
Low Grade	163	20	183	
High Grade	68	55	123	
Total	231	75	306	
Finasteride				
Prostatectomy				
Low Grade	93	17	110	
High Grade	39	76	115	
Total	132	93	225	

**Table 5**

Characteristics of participants with prostate cancer who had a prostatectomy

	No Prostatectomy (N=1484)	Prostatectomy (N=531)	P-value <sup>a</sup>
Finasteride Arm	596 (40.2%)	225 (42.4%)	0.42
Baseline PSA	1.5 (0.9, 2.1) <sup>b</sup>	1.6 (1.0, 2.3)	0.002
Baseline Age	65 (60, 69)	61 (58, 64)	< 0.0001
Race			0.10
White	1372 (92.5%)	496 (93.4%)	
African American	79 (5.3%)	18 (3.3%)	
Other	33 (2.2%)	17 (3.2%)	
Family History of Prostate Cancer	312 (21.0%)	122 (22.9%)	0.38
Referred to Biopsy because of PSA	336 (22.6%)	171 (32.2%)	< 0.0001
Referred to Biopsy because of DRE	265 (17.9%)	139 (26.2%)	< 0.0001
Prior Negative Biopsy	205 (13.8%)	68 (12.8%)	0.61
High Grade Cancer from Biopsy	395 (26.6%)	168 (31.6%)	0.03

<sup>a</sup> Chi-square test for categorical variables, Wilcoxon rank sum test for continuous covariates (baseline PSA and age).

<sup>b</sup> Median (1st quartile, 3rd quartile) reported for continuous covariates.