Estimating Rates of True High-Grade Disease in the Prostate Cancer Prevention Trial

Paul Pinsky, Howard Parnes, and Leslie Ford

Abstract

The Prostate Cancer Prevention Trial (PCPT) showed a decreased prostate cancer rate but an increased rate of high Gleason grade disease on biopsy for finasteride versus placebo. The results from radical prostatectomy (RP) on 25% of the men undergoing RP have recently been reported and suggest that grading artifacts in biopsy Gleason scoring may have occurred. We used a statistical model to extrapolate the RP Gleason results to all men in the PCPT using a missing-at-random assumption. We estimated the rates of true high-grade (Gleason 7-10) and true low-grade disease, where true Gleason grade is what is (or would have been) found on RP. We also estimated misclassification rates on biopsy of true high-grade and low-grade disease. We show that the rate of upgrading of biopsy low-grade disease to high-grade on RP is a function of misclassification rates as well as the ratio of true low-grade to high-grade disease. The estimated relative risks for true low-grade and true high-grade disease for finasteride compared with placebo were 0.61 (95% confidence interval, 0.51-0.71) and 0.84 (95% confidence interval, 0.68-1.05), respectively. The misclassification rate of true high-grade disease (to low-grade disease on biopsy) was significantly lower for finasteride (34.6%) than for placebo (52.6%). Although misclassification rates differed, upgrading rates were similar in each arm due to the different ratios of true low-grade to high-grade disease in each arm. Results from RP show that misclassification rates on biopsy were higher in the placebo arm and that the rate of true high-grade disease may have been lower in the finasteride arm.

The Prostate Cancer Prevention Trial (PCPT), a randomized trial of the 5α-reductase inhibitor finasteride versus placebo, was recently completed and showed an overall decreased relative risk (RR) for the cumulative prevalence of prostate cancer (1). However, the RR for high-grade disease, as defined by biopsy Gleason score of 7-10, was elevated in the finasteride arm. The relevance of this finding, however, has been called into question by concerns that the biopsy Gleason score may be affected by volume or other artifacts induced by exposure to finasteride. Because finasteride is known to shrink the prostate, it has been argued that biopsy Gleason scores may be more accurate in the finasteride arm than in the placebo arm because a greater proportion of the prostate is being sampled (2, 3). Accordingly, there may be less upgrading of Gleason score from biopsy to radical prostatectomy (RP) in men on finasteride. This effect could lead to differential scoring between arms and an artificially higher rate of high-grade disease on biopsy in the finasteride arm.

Recently, the results of pathology specimens from ~25% of men in the PCPT with prostate cancer who underwent RP were reported (4). The researchers concluded that the effects of finasteride on prostate volume may have introduced detection bias (i.e., increased detection of high-grade cancers in the smaller, finasteride-treated glands). However, the observation that the rate of upgrading (from low grade at biopsy to high grade at RP) was similar in each arm was somewhat paradoxical.

Whereas the analyses published to date shed much light on the possible reasons for the apparent increase in high-grade disease on finasteride, they have not directly answered all the important questions about finasteride use and have not provided a quantitative estimate of the “true” RR for high-grade disease. In addition, although the original report of the RP results stated that upgrading and downgrading rates were dependent on the prevalence of high-grade disease at prostatectomy in each arm, it did not fully explain the relationship between differences (or lack thereof) in upgrading rates between arms and differences in high-grade and low-grade disease prevalence, nor did it derive a quantitative relationship between the two.

The purpose of this article is to extrapolate the findings from the RP study to the overall population of men in the PCPT with prostate cancer to estimate the overall rates of true Gleason high-grade (and low-grade) disease in each arm, where the true Gleason grade is what was, or would have been true at RP. We also estimated misclassification rates on biopsy of true high-grade and low-grade disease, where true Gleason grade is what would have occurred. We used a statistical model to extrapolate the RP Gleason results to all men in the PCPT using a missing-at-random assumption. We estimated the rates of true high-grade (Gleason 7-10) and true low-grade disease, where true Gleason grade is what was, or would have been found on RP. We also estimated misclassification rates on biopsy of true high-grade and low-grade disease. We show that the rate of upgrading of biopsy low-grade disease to high-grade on RP is a function of misclassification rates as well as the ratio of true low-grade to high-grade disease. The estimated relative risks for true low-grade and true high-grade disease for finasteride compared with placebo were 0.61 (95% confidence interval, 0.51-0.71) and 0.84 (95% confidence interval, 0.68-1.05), respectively. The misclassification rate of true high-grade disease (to low-grade disease on biopsy) was significantly lower for finasteride (34.6%) than for placebo (52.6%). Although misclassification rates differed, upgrading rates were similar in each arm due to the different ratios of true low-grade to high-grade disease in each arm. Results from RP show that misclassification rates on biopsy were higher in the placebo arm and that the rate of true high-grade disease may have been lower in the finasteride arm.

Authors’ Affiliations: Division of Cancer Prevention, National Cancer Institute, Bethesda, Maryland.

Received 12/17/2007; revised 04/23/2008; accepted 04/23/2008.

Requests for reprints: Paul F. Pinsky, Division of Cancer Prevention, National Cancer Institute, 6130 Executive Boulevard, EPN 3064, Bethesda, MD 20892.

E-mail: pp4f@nih.gov.

©2008 American Association for Cancer Research.

doi:10.1158/1940-6207.CAPR-07-0007
been, found on RP. As part of this exercise, we also show that the misclassification rates, defined here as the proportion of men who have high-grade (low-grade) disease on RP and low-grade (high-grade) disease on biopsy, are a better measure of the possible grading artifact induced by finasteride than the proportion upgraded. As we will show, the proportion upgraded (or downgraded) in each study arm from biopsy to RP is a function of both the misclassification rates as defined above and the relative distribution of true (i.e., on RP) high-grade versus low-grade disease in each arm. This point is critical in the PCPT because the relative distribution of true high-grade versus low-grade disease may, and we estimate does, differ between study arms.

**Materials and Methods**

**The PCPT**

The PCPT randomized 9,423 men to finasteride and 9,459 to placebo (1). Subjects were followed over a 7-y period, during which they were monitored for compliance and were screened annually with prostate-specific antigen (PSA) and digital rectal exam. Subjects with an elevated PSA or suspicious digital rectal exam were referred for biopsy. In addition, all subjects without cancer diagnosis at the end of the study were asked to return for an end-of-study biopsy. Based on a cutoff date of the unblinding of the PCPT (on 6/23/2003) or of participants’ end-of-study window (90 d after the 7-y anniversary), whichever came first, a total of 823 and 1,194 cases of prostate cancer were reported in the finasteride and placebo arms, respectively (Table 1). The RRs for biopsy high-grade (7-10) and low-grade (2-6) disease for finasteride versus placebo were 1.15 and 0.56, respectively, with an overall RR of 0.69. The results of the RP study are summarized in Table 2 (4). About 25% of men with prostate cancer in each arm underwent RP. Of those with RP, ~75% of subjects in each arm with high-grade disease on biopsy (Gleason 7-10) remained high-grade on RP, and a similar percentage of subjects with low-grade disease on biopsy (Gleason 2-6) remained low-grade on RP.

**Statistical Methods**

We take the gold standard for Gleason grade to be the RP results. Statistically, the problem then becomes one of missing data, specifically, how to impute the gold standard result to those men who did not have the gold standard test of RP done. We modeled these data assuming a missing at random mechanism, which in this context means that the probability of undergoing RP, given the biopsy Gleason score category (2-6, 7, 8-10) and study arm, is independent of the true Gleason score category (5). This implies that within each study arm and biopsy Gleason category, those undergoing RP have the same probability distribution of true Gleason categories as those not undergoing RP. The model attempts to estimate in each arm the rates of true high-grade and low-grade disease, as well as the corresponding misclassification rates on biopsy of the true Gleason category. The model was fit using maximum likelihood. Further details of the model and the estimation methods are given in Appendix A.

**Misclassification and Upgrading**

The misclassification rates, along with the relative overall rates of true high-grade and low-grade disease in each arm, determine the rate of “upgrading” from biopsy to RP in each arm, as we show below. The rate of low-grade disease on biopsy [in each arm; LG(Bx)] will be the rate of true low-grade disease times the probability that there is no misclassification on biopsy, plus the rate of true high-grade disease times the probability that there is misclassification on biopsy. Thus, LG(Bx) rate = TLG (1 - MLG) + THG MHG, where TLG and THG are the rates of true low-grade and high-grade disease, respectively, and MLG and MHG are the misclassification rates (on biopsy) of true high-grade and low-grade disease.
Results

Table 3 displays the rates of true high-grade (7-10) and low-grade (2-6) disease in each arm estimated by the model. Rates of true high-grade disease were estimated at 3.9% for finasteride and 4.6% for placebo, for a RR of 0.84 [95% confidence interval (95% CI), 0.68-1.05]. For true low-grade disease, estimated rates were 4.9% (finasteride) and 8.0% (placebo) for a RR of 0.61 (95% CI, 0.51-0.71). The RR of true high-grade disease was lower in the finasteride arm (4.9/3.9 = 1.26) as compared with the placebo arm (52.6%). Among the true high-grade cases, misclassification to low grade on biopsy was relatively rare for true 8-10 cases (6% and 10.6% for finasteride and placebo, respectively) but significantly more common for true 7 cases (41.6% for finisteride and 58.3% for placebo). Misclassification of true low-grade disease to high-grade on biopsy was slightly (and borderline statistically significantly) higher in the finasteride (15.2%) than in the placebo arm (8.8%; P = 0.05).

The estimated rates of true high-grade and low-grade disease, as well as the estimated misclassification rates, explain the similar rates of upgrading in each arm. Figure 1 shows a schematic that breaks down all of the cancers in each arm into true status by biopsy status (low grade or high grade) according to the model estimates. As shown in Methods, the proportion upgraded depends not only on the misclassification rate of true high-grade disease but also on the ratio of true low-grade to high-grade disease, as well as the misclassification rate of true low-grade disease. Although the misclassification rate of true high-grade disease was lower in the finasteride arm, the ratio of true low-grade to high-grade disease was also lower in the finasteride arm (4.9/3.9 = 1.26) as compared with the placebo arm (8.0/4.6 = 1.74). Thus, these two effects essentially canceled each other out, leaving a similar proportion of upgrading from biopsy in each arm.

Discussion

We have used a statistical model here to extrapolate the RP results in the PCPT to the entire population of men with prostate cancer. The model allowed us to estimate the RR for true
Based on only the observed data from Table 2, it is seen that 30% (27 of 89) of true high-grade cancers were misclassified as low grade on biopsy in the placebo arm, as compared with 50% (52 of 105) in the placebo arm. The estimated misclassification rates derived here, 34.6% and 52.6%, respectively, are close to those figures. However, these raw, observed misclassification rates are biased estimates of the true rates because they depend on the relative proportions of men in each biopsy Gleason category who receive RP. For example, if the RP rate for the biopsy Gleason 2-6 placebo-arm men was twice of what it was, then all other things being equal, one would have expected 104 men with biopsy Gleason 2-6 and RP Gleason 7-10 instead of 52, and the resulting observed rate of misclassification would have been 66% (104 of 157) instead of 50%. The modeled estimate would still be 52.6%, however. The raw estimates turn out to be close to the modeled estimates only because the rate of undergoing RP did not differ that greatly across the biopsy Gleason categories.

This analysis takes prostate volume into account (indirectly) to the extent that it, and possibly other factors, may affect the relative validity of the biopsy Gleason scoring in the finasteride versus placebo arms. However, it does not take into account the fact that, because of volume effects, more cancers (presumably of all grades) would likely be missed on biopsy in the placebo than in the finasteride arm. Taking this into account could further decrease the RR for true high-grade disease. A recent statistical analysis of prostate volume in the PCPT showed a RR of 0.79 for high-grade cancer versus no cancer associated with a 10-cm³ increase in gland volume (6). Given the average difference in gland volume cited in that study of ~9 cm³, this would translate into an average RR of 0.82 for the detection of high-grade cancer for placebo versus finasteride. Correcting for this detection bias effect would then reduce the RR for high-grade disease for finasteride versus placebo by a factor of 0.82. Thus, the estimates presented here are likely to be conservative in terms of the protective effect of finasteride.

The main limitation of this analysis is its reliance on the missing at random assumption (i.e., that the decision to undergo RP, given study arm and biopsy Gleason, did not depend on the true Gleason score). This assumption could be violated if there were other factors, besides study arm and biopsy Gleason, that influenced the decision to undergo RP and that were also related to true Gleason status. Two observed factors that were associated with having RP were age and PSA (4). Men receiving RP were ~4 years older on average, in each arm, than men not receiving RP, and in the finasteride arm, men receiving RP had higher median PSA (3.5) than those not receiving RP; median PSA in the control-arm men was similar among those who did and did not receive RP.

If age and PSA were also related to true Gleason score, conditional on biopsy Gleason, then our original results could be biased. Therefore, to examine the possible effect of these covariates, we expanded our model to include the effects of age and PSA. Specifically, in each arm, we allowed the probability of true Gleason disease to depend on age and PSA (both dichotomized as below or above the median level). The RR estimates derived from this model were essentially the same as those derived from the simpler model: For true high Gleason grade disease, the RR from the expanded model was 0.83 compared with the original RR of 0.84.

There may be other unmeasured factors that affected both RP status and true Gleason status, and it is thus possible that these could be introducing a bias into our estimates. Note, however, that if an unmeasured factor were similarly
associated in each arm with undergoing RP and with true Gleason score, then controlling for that factor could change the estimates of the rate of high-grade disease in each arm, but would tend to have a much smaller effect on the estimate of the RR of finasteride versus placebo, because the disease rate estimates for both arms would likely be affected in a similar manner. This was the case with the factors of age and PSA examined above, where, although the RR estimate for true high Gleason grade disease was basically unchanged, the estimates of the rates of true high-grade disease in each arm did increase modestly (from 3.9 to 4.1 in the finasteride arm and from 4.6 to 4.9 in the placebo arm).

In conclusion, results from the analysis of RP specimens in the PCPT show that misclassification rates on biopsy were indeed higher in the placebo than in the finasteride arm and that the rate of true high-grade disease may have been lower in the finasteride arm. In addition, the upgrading paradox (similar upgrading rates in the two arms despite less misclassification on biopsy in the finasteride arm) is explained by the fact that finasteride decreased the ratio of true low-grade to high-grade disease, resulting in more than the expected rate of upgrading on this arm.

**Disclosure of Potential Conflicts of Interest**

No potential conflicts of interest were disclosed.

**References**

Estimating Rates of True High-Grade Disease in the Prostate Cancer Prevention Trial
Paul Pinsky, Howard Parnes and Leslie Ford
Cancer Prev Res Published OnlineFirst May 18, 2008.