EBRU Session 5

Methodological Review by Jeffrey D. Blume, PhD

Article Reviewed


Review

This article reports on a clinical trial comparing dutasteride to placebo for reducing the incidence of prostate cancer in men who are at increased risk for prostate cancer. This trial is an example of a well-conducted and well-controlled study that, unfortunately, yields suspect conclusions because of an unsophisticated and poorly executed statistical analysis.

Are the results valid?

There are three things we need to consider: internal validity, external validity, and the quality of the statistical analysis.

Internal validity addresses the reproducibility of the study conduct and/or its implementation. At issue is the standardization of, and adherence to, recruitment and treatment procedures, data capture and cleaning procedures, and the reliability of endpoint assessments. Ideally, the study procedures should be sufficiently regularized so that if the study were repeated again, on the same set of participants, we would observe the same data. From reading the paper and supplementary appendix, it appears the study was well designed and well conducted. The discussion of the concordance of local and central pathology is a good example of this. However, the paper does not detail protocol violations and protocol deviations, so we really don’t get a good sense of how well the sites and enrolling physicians adhered to the protocol.

External validity addresses generalizability. The external validity of a study is largely determined by the characteristics of the participant population, characteristics of the treating physicians, and the treatment and follow-up procedures. A necessary condition for high external validity is high internal validity. However, this is not a sufficient condition. Well-conducted studies often do not generalize because, for example, the study procedures are too laborious or not suitable for everyday clinical practice. In this study, the results would generalize to a clinical population with the characteristics given in Table 1 (total column) who are willing to enroll in a clinical trial (it has been shown that people willing to enroll in a trial are generally healthier than the general population). So for example, these results pertain to (slightly healthier) white males. There were too few Black, Asian, and Hispanic participants to reliably learn anything about them.

Next we turn to the statistical analysis. Without a proper statistical analysis, data cannot be generalized. In fact, most underlying statistical models are nothing more than mathematical descriptions of the generalized population. Thus, it is important to understand the assumptions and the applicability of statistical models, as this is the formal mechanism through which data are made scientifically interpretable and generalizable.

Unfortunately, there were a number of problems with the statistical analysis for this study. The primary endpoint is stated as the time to biopsy-detectable prostate cancer. But because the endpoint is assessed only twice – once at 24 months and once at 48 months – it does not make sense to treat this
EBRU Session 5

variable as if it were continuous. The paper seems confused on this point, as it reports reductions in relative risk (which can only come from a binary endpoint) using a 95% CI, yet at the same time provides a p-value from a log-rank test (Mantel-Cox test) which is used to compare two empirical time-to-event curves and where the proper metric is a hazard ratio, not a relative risk. Also, the text does not distinguish between an intention-to-treat analysis versus an as-treated analysis, so it is not clear which is performed here.

There is a large amount of missing data that is neither discussed nor accounted for. Figure 1 indicates that 17% of the primary endpoint data is missing. In cases like this where there is a large amount of data missing, the authors are obligated to show that the results are insensitive to the missing data. This can be done with a sensitivity analysis, or better yet, an analysis that explicitly adjusts for the missing data (e.g., with multiple imputation or a pattern mixture model). The supplementary appendix suggests two approaches for handling missing data: complete case analysis and last observation carried forward (LOCF). Both approaches should be avoided, although it appears that the complete case analysis approach was used for this paper.

Why should we care about the missing data? Why not just ignore it? Because if the missing biopsy data were not missing completely at random, then the results based on the complete case analysis will be biased and could even have been driven by the differential missingness. The amount of missing data, especially for the primary endpoint, is a good surrogate marker for trial quality. I typically don’t worry about missing data if it comprises less than 5% of the data available for the intention-to-treat analysis set. Interestingly, the log-rank test should be able to accommodate participants without biopsy information by censoring them. So it is not entirely clear why these participants were excluded in the first place.

The analysis also omits a multivariate model to adjust for observed covariate patterns. In this case, adjusting would mean using a Cox model (for the time-to-event endpoint) or a logistic regression model (for a binary endpoint). Table 3 is a crude attempt to adjust for Gleason score. It would have been much easier, more powerful, more reliable to use a multivariate model to do this. Also, a critical omission is a multivariate model that adjusts for site differences. This is especially important when sites come from different countries, as is the case here. Such models are the standard for large multi-center trials, because site variation can mask, or create the illusion of, treatment effects when not properly accounted for.

Side note: Randomization only promotes balance among treatment groups. A randomized trial is not necessarily a balanced trial. It is balance that is essential for fair and valid scientific comparisons. Thus, trial results should always be adjusted for imbalances in important pre-specified covariates, i.e., one should always use a multivariate model. I typically report both the unadjusted (univariate) and adjusted (multivariate) comparisons, as both are important for understanding what the data say.

What are the results?

Due to the limitations of the statistical analysis above, I would say it remains unclear what the results are.

Can the results be applied to patient care?

This is not immediately clear due to the concerns raised above.
EBRU Session 5

References

Two excellent references on the fundamentals of clinical trials are listed below. The text by Piantadosi is a comprehensive text that is excellent for quick referencing. The Friedman et al. text has less jargon, but is not as comprehensive.

Clinical Review by Daniel A. Barocas, MD, MPH

Article Reviewed

Review
This is a multi-center, randomized, placebo-controlled trial in which dutasteride 0.5 mg daily was compared to placebo in men ages 50-75, with a PSA between 2.5 and 10 ng/mL and a prior negative prostate biopsy. The aim was to determine whether dutasteride reduces the risk of diagnosis of prostate cancer at scheduled biopsies 2 and 4 years after starting treatment. Among 6,729 men who underwent at least one biopsy, the absolute risk reduction with dutasteride was about 5% and the relative risk reduction was about 23% (25.1% vs. 19.9%).

Are the results valid?
"Yes, but..." This is a very professionally designed and executed study and I do not think one can question its reliability or the precision with which the results reflect the effect of the drug on the measured endpoints in this study population. The study was adequately powered, met accrual goals and identified an effect size that was anticipated, has narrow confidence intervals, is relatively consistent across strata of baseline characteristics and is similar to a prior study of finasteride for prostate cancer prevention (PCPT).

Validity speaks to the study's generalizability or usefulness in clinical scenarios. The question we would all like to answer is whether dutasteride reduces the risk of important prostate cancers (i.e., clinically-detected cancers, and particularly those with the potential to cause morbidity and/or mortality) sufficiently to justify its cost and risks. This would inform the patients and clinicians about whether to use dutasteride for prostate cancer prevention in the real-world setting. The REDUCE trial set out to answer a slightly different question: the primary endpoint was the number of incident prostate cancers after 2 or 4 years of treatment, in patients undergoing protocol-mandated biopsies every 2 years. This broadens the focus to include cancers that may not have been detected otherwise. The investigators found that much of the reduction in cancer incidence was in the moderately well-differentiated cancers (Gleason 5 or 6), which were found in 437 (13.2%) in the dutasteride group and 617 (18.1%) in the placebo group. While this translated into an overall reduction in the incidence of prostate cancer in the treatment group, there was no difference in incidence between the placebo and treatment groups on protocol-independent biopsies (16.7% vs. 16.6%). Thus, it is not clear that the use of dutasteride would result in a reduction in prostate cancer incidence in a clinical setting.

What are the results?
There are three main findings presented in this paper: 1) Dutasteride was associated with a reduction in prostate cancer incidence; 2) Most of the effect was driven by a reduction in Gleason 5 and 6 cancers in the protocol-mandated biopsies; 3) Dutasteride was associated with salutary effects on urinary symptoms and BPH-related outcomes, but was also associated with more sexual side effects.

Dutasteride resulted in an absolute reduction in the incidence of prostate cancer of 5.1% and a relative reduction of 23%. This reduction in risk held up across strata of baseline variables (age, family history, International Prostate Symptom Score [IPSS], prostate volume, PSA level and body-mass index) and over time (years 1 and 2 vs. years 3 and 4). The effect size is remarkably similar to that identified in the PCPT.
Again, this reduction in risk really seems to be limited to the Gleason 5 and 6 cancers and was not evident in the protocol-mandated (‘for-cause’) biopsies. Looking at Gleason score, there was a slightly higher proportion of Gleason 7 to 10 cancers in the treatment group than the placebo group, both on protocol-independent biopsies (7.1% vs. 5.6%) and all biopsies performed (33.5% vs. 27.4%). This difference is attenuated compared to the discrepancy seen in the PCPT and could simply reflect the reduced number of Gleason 5 and 6 cancers in the treatment group. However, there was a concerning finding in years 3 and 4, where 12 Gleason 8-10 cancers were diagnosed in the dutasteride group compared to 1 in the placebo group (p=0.003). Thus, despite the strength of the study and reliability of its findings, it does not conclusively demonstrate a reduction in the incidence of clinically-apparent disease and it does not resolve public concerns about the possibility of increased risk of high-grade prostate cancer with chronic 5-alpha-reductase inhibition.

There was clear improvement in each of the non-cancer, voiding-related outcomes in the dutasteride arm. Men on treatment had a 40.7% lower incidence of urinary tract infection (5.3% vs. 8.8%), a 77.3% lower incidence of acute urinary retention (1.6% vs. 6.7%), and 73.0% fewer BPH-related surgeries (1.4% vs. 5.1%) compared to men on placebo. This corresponded to a mean reduction in prostate volume of 17.5% in the treatment group compared to an increase of 19.7% in the placebo group. Lastly, in men with moderate to severe baseline lower urinary tract symptoms, there was more improvement in IPSS score in the dutasteride group (decrease of 3.9 points vs. 1.3 points). However, sexual function side-effects, such as low libido, erectile dysfunction and decreased semen volume, were more common among men taking dutasteride.

Can the results be applied to patient care?

The concerns about widespread utilization of long-term dutasteride administration for prostate cancer prevention are 1) the absence of an obvious benefit in prevention of clinically-detected cancers, and; 2) that there may be a predilection for high-grade cancers in men treated with 5-alpha reductase inhibitors. In the final analysis, the side effects, cost of long-term treatment [estimated at $1500/year] [Walsh editorial], uncertain benefits in a general population and concerns about high-grade cancer will limit the wide-scale use of dutasteride.4 This trial was designed to evaluate dutasteride’s efficacy in a high-risk population, though the risk level of these patients (with a prior negative biopsy and a PSA between 2.5 and 10) is up for some debate. There may be a truly high-risk group for whom the benefits of chemo-prevention with dutasteride outweigh the costs and risks. Nonetheless, we probably will continue to see use of dutasteride in the setting in which it was tested – patients with elevated PSA and a prior negative biopsy, who represent an anxiety-provoking clinical management challenge – based perhaps more on the uncertainty of the clinical situation than on the conclusions of this study. Finally, for patients with BPH-related indications for 5-alpha reductase inhibition, the balance of risks and benefits seems to favor use of dutasteride.

References