APPENDIX A: Methods for Evidence Synthesis

Literature Search

To identify relevant citations, we searched Ovid MEDLINE (1966 to July 2006.) For Key Question #1 we used the following search strategy:

1. exp Adrenergic alpha-Antagonists/ad, ae, cl, tu, ct, du [Administration & Dosage, Adverse Effects, Classification, Therapeutic Use, Contraindications, Diagnostic Use]
2. exp Prostatic Hyperplasia/mo, cl, co, di, pc, dh, dt, ep, su, th, ge [Mortality, Classification, Complications, Diagnosis, Prevention & Control, Diet Therapy, Drug Therapy, Epidemiology, Surgery, Therapy, Genetics]
3. 1 and 2
4. limit 3 (humans and male and "all adult (19 plus years)" and (clinical trial, phase i or clinical trial, phase ii or clinical trial, phase iii or clinical trial, phase iv or clinical trial or controlled clinical trial or evaluation studies or multicenter study or randomized controlled trial or technical report))
5. explode Finasteride
6. 3 and 5

For Key Questions #2 and #3, we used steps 1 to 4 of the same search string. We searched the Cochrane Database of Systematic Reviews (2nd quarter, 2006) but did not identify any additional systematic reviews.

All citations were imported into an electronic database (EndNote 9.0).

Study Selection

One reviewer assessed abstracts of citations identified from literature searches for inclusion, using the criteria described above. Full-text articles of potentially relevant abstracts were retrieved and a second review for inclusion was conducted by reapplying the inclusion criteria.

Data Abstraction

The following data were abstracted from included trials: study design, setting, population characteristics (including sex, age, ethnicity, diagnosis), eligibility and exclusion criteria, interventions (dose and duration), comparisons, numbers screened, eligible, enrolled, and lost to follow-up, method of outcome ascertainment, and results for each outcome. We recorded intention-to-treat results when reported. In cases where only per-protocol results were reported, we calculated intention-to-treat results if the data for these calculations were available. In trials with crossover, outcomes for the first intervention were recorded if available. This was because of the potential for differential withdrawal prior to crossover biasing subsequent results and the possibility of either a “carryover effect” (from the first treatment) in studies without a washout period, or “rebound” effect from withdrawal of the first intervention.

Data abstracted from observational studies included design, eligibility criteria duration, interventions, concomitant medication, assessment techniques, age, gender, ethnicity, number of patients screened, eligible, enrolled, withdrawn, or lost to follow-up, number analyzed, and results.

Quality Assessment

We assessed the internal validity (quality) of trials based on the predefined criteria listed in Appendix B. These criteria are based on the U.S. Preventive Services Task Force and the National Health Service Centre for Reviews and Dissemination (U.K.) criteria.[22, 23] We rated the internal validity of each trial based on the methods used for randomization, allocation concealment, and blinding; the similarity of compared groups at baseline; maintenance of comparable groups; adequate reporting of dropouts, attrition, crossover, adherence, and contamination; loss to follow-up; and the use of intention-to-treat analysis. Trials that had a fatal flaw in one or more categories were rated "poor-quality"; trials that met all criteria were rated "good-quality"; the remainder were rated "fair-quality." A fatal flaw occurs when there is evidence of bias or confounding in the trial, for example when randomization and concealment of allocation of random order are not reported and baseline characteristics differ significantly between the groups. In this case, randomization has apparently failed and for one reason or another bias has been introduced.

As the fair-quality category is broad, studies with this rating vary in their strengths and weaknesses: the results of some fair-quality studies are *likely* to be valid, while others are only *probably* valid. Those studies considered only *probably* valid are indicated as such using a "fair-poor" rating. A poor-quality trial is not valid—the results are at least as likely to reflect flaws in the study design as the true difference between the compared drugs. External validity of trials was assessed based on whether the publication adequately described the study population, how similar patients were to the target population in whom the intervention will be applied, and whether the treatment received by the control group was reasonably representative of standard practice. We also recorded the role of the funding source.

Appendix B also shows the criteria we used to rate observational studies. These criteria reflect aspects of the study design that are particularly important for assessing adverse event rates. We rated observational studies as good-quality for adverse event assessment if they adequately met six or more of the seven predefined criteria, fair-quality if they met three to five criteria and poor-quality if they met two or fewer criteria.

Included systematic reviews were also rated for quality based on pre-defined criteria (see Appendix B), based on a clear statement of the questions(s), inclusion criteria, adequacy of search strategy, validity assessment and adequacy of detail provided for included studies, and appropriateness of the methods of synthesis.

Overall quality ratings for the individual study were based on internal and external validity ratings for that trial. A particular randomized trial might receive two different ratings: one for effectiveness and another for adverse events. The overall strength of evidence for a particular key question reflects the quality, consistency, and power of the set of studies relevant to the question.

Evidence Synthesis

An evidence report pays particular attention to the generalizability of efficacy studies performed in controlled or academic settings. Efficacy studies provide the best information about how a drug performs in a controlled setting that allow for better control over potential confounding factors and bias. However, efficacy studies have some limitations, as the results are not always applicable to many, or to most, patients seen in everyday practice. This is because most efficacy studies use strict eligibility criteria which may exclude patients based on their age, sex, medication compliance, or severity of illness. For many drug classes severely impaired patients are often excluded from trials. Often, efficacy studies also exclude patients who have "comorbid" diseases, meaning diseases other than the one under study. Efficacy studies may also use dosing regimens and follow up protocols that may be impractical in other practice settings. They often restrict options, such as combining therapies or switching drugs, that are of value in actual practice. They often examine the short-term effects of drugs that, in practice, are used for much longer periods of time. Finally, they tend to use objective measures of effect that do not capture all of the benefits and harms of a drug or do not reflect the outcomes that are most important to patients and their families.

Data Presentation

We constructed evidence tables showing the study characteristics, quality ratings, and results for all included studies. Studies that evaluated one macrolide against another provided direct evidence of comparative benefits and harms. Outcomes of changes in symptom measured using scales or tools with good validity and reliability are preferred over scales or tools with low validity/reliability or no reports of validity/reliability testing. Where possible, head-to-head data are the primary focus of the synthesis. No meta-analyses were conducted in this review due to heterogeneity in treatment regimens, use of concomitant medications, outcome reporting and patient populations.

In theory, trials that compare these drugs to other interventions or placebos can also provide evidence about effectiveness. This is known as an indirect comparison and can be difficult to interpret for a number of reasons, primarily issues of heterogeneity between trial populations, interventions, and assessment of outcomes. Indirect data are used to support direct comparisons, where they exist, and are also used as the primary comparison where no direct comparisons exist. Such indirect comparisons should be interpreted with caution.

Appendix B: Trials comparing alpha antagonists

Clinical Trial	Inclusion Criteria/Pt. Population	Intervention	Results	Safety/Comments
Cam et al 2003 Prospective clinical study 178 patients doxazosin 4 mg 24 months no financial disclosure Turkish	Men > 50 years old IPSS 18-35 Attendance to a urology department due to: -LUTS -age > 50 -unremarkable medical hx in terms of LUTS -no definitive need for surgery	Doxazosin 1 mg titrated to 4 mg	Reduction in symptom scores from 24 (SD ±7.4) before medication to 17 (SD ±8.4) after 3 months of treatment In the patients reporting doxazosin as ineffective, no change, or effective, 93%, 59% and 15% respectively underwent surgery. Of the 178 patients enrolled 47% underwent surgery.	Evaluation of the efficacy doxazosin was determined by one multiple choice question regarding the satisfaction with the medical treatment in terms of relieving symptoms
Ichioka et al 2004 Prospective 123 patients 43 months Tamsulosin (n = 123) No financial disclosure Japanese	Men 53-88 years old Dx BPH Treated with tamsulosin >12 months	Tamsulosin 0.2 mg titrated up to 0.4 mg as needed to relieve sx.	Predictive for treatment failure: baseline IPSS ≥ 15, months 0-12 lowest IPSS ≥ 13, lowest QoL score of ≥ 3 and lowest BPH impact score of ≥ 4.	
Roehrborn et al 1996 Prospective, placebo controlled, randomized, double-blinded 2084 patients 1 year (Terazosin n = 1053 Placebo n = 1031) Funding: Abbot Labratorie American	Men ≥ 55 years old Moderate-severe BPH AUA-Symptom Score (SS) ≥ 13 AUA-Bother Score (BS) ≥ 8 PUF ≤ 15 mL/sec. Voided volume = 150 mL	Terazosin 1 mg x 3 days, 2 mg x 25 days, ↑ 5 mg or 10 mg as tolerated Placebo	Statistically superior improvements were observed in regard to AUA-BS, BPH impact index and the QoL score in terazosin-treated patients. PUF improved Treatment failure was higher in placebo	Withdraw was higher due to ADR was higher in terazosin patients

Okada et al 2000 Single-blind, randomized 61 patients 4 weeks Japanese	Symptomatic BPH	Terazosin 1-2 mg Tamsulosin 0.2 mg	Both meds significantly improved the total IPSS, irritative and obstructive symptom score, and quality of life. There was no significant difference for these variables between groups. There was no significant improvements between groups.	Incidence of ADR was not significantly different between groups. Neither medication affected systolic or diastolic blood pressure.
Lee et al 1997 Single-blind 98 patients randomized 8 weeks Korean	Moderate to severe BPH	Tamsulosin 0.2 mg Terazosin 1 mg ↑ 5 mg	Both medications similarly improved IPSS and Increased Qmax	Terazosin: -systolic and diastolic BP decreased significantly -dizziness, dry mouth were more frequent
Tsuiji et al 2000 Open-label 105 patietns Randomized (Prazosin n= 32, Terazosin n=35, Tamsulosin n=38) 4 weeks	LUTS associated with BPH	Prazosin 1 mg ↑ 2 mg Terazosin 1 mg ↑ 2 mg Tamsulosin 0.1 mg ↑ 0.2 mg	All significantly reduced subjective symptom scores from baseline. Terazosin significantly better improvement than Tamsulosin in 4 of 9 symptom scores (urgency, sense of residual urine, prolonged micturition, intermittency) Significant increase in flow with prazosin	ADR which lead to withdrawal: Prazosin = 1 Terazosin = 3 Tamsulosin = 0
Na et al 1998 Single-blinded, randomized 212 Patients Randomized 4 weeks Chinese	BPH	Terazosin 2 mg Tamsulosin 0.2 mg	Tamsulosin and terazosin: significant improvement in IPSS, Qmax and average urinary flow rate from baseline	Dizziness, hypotension occurred significantly more frequently with terazosin than tamsulosin