ELSEVIER

Urologic Oncology: Seminars and Original Investigations 29 (2011) 462–466

UROLOGIC ONCOLOGY

Seminar article

Advanced topics in evidence-based urological oncology: Using results of a subgroup analysis☆

Charles D. Scales Jr., M.D.[a], Steven E. Canfield, M.D.[b,*]

[a] *Division of Urologic Surgery, Duke University Medical Center, Durham, NC 27710, USA*
[b] *Department of Surgery, Division of Urology, University of Texas Health Science Center, Houston, TX 77030, USA*

Received 4 November 2010; received in revised form 4 January 2011; accepted 9 January 2011

Abstract

Background: Randomized, controlled trials are the cornerstone of evaluation for therapies in urologic oncology. Clinicians and investigators are frequently interested in whether treatment effects differ among subgroups of patients defined by clinically relevant characteristics.

Methods: The evidence-based approach to subgroup analysis is explored using an example from the urologic literature. Potential reasons why the results of subgroup analyses may not be reliable are reviewed. Criteria for assessing the validity of a subgroup effect are described.

Results: The likelihood of observing clinically important differences in treatment results by chance increases with each additional comparison of groups. Ideally, subgroup comparisons should be specified a priori and few in number. The probability of observing the difference in outcomes due to chance should be low and, ideally, the difference will be large. Finally, external evidence or biologic data should support the hypothesized difference in subgroup outcomes.

Conclusion: Use of these criteria for subgroup analyses will promote a more evidence-based management for oncologic diseases within urology. Understanding appropriate use of subgroup analyses will help clinicians target therapies towards those patients most likely to benefit, and avoid both limiting potentially beneficial therapies or utilizing ineffective therapies when observed subgroup treatment effects are likely due to chance.

Keywords: Subgroup analysis; Radical prostatectomy; Watchful waiting

1. Introduction

Urologists seek the best evidence to guide decision-making when treating patients. Ideally, the evidence for therapeutic decisions derives from randomized controlled trials, which in the absence of meta-analyses, generate results that are least susceptible to bias and constitute the highest level of evidence [1]. However, the results of an RCT are typically presented as an average effect across the study population; clinicians ideally want more specific information to assist them in applying trial results to individual patients. Thus, clinicians may be acutely interested in treatment effects among subgroups of trial patients, such as male vs. female or older vs. younger patients.

Investigators reflect this interest in clinically important groups by frequently including subgroup analyses in the reports of clinical trials. For example, over half of clinical trial reports in 3 leading medical journals include at least 1 subgroup analysis [2]. However, the practice of subgroup analysis frequently provides misleading results, typically because of the increased risk of false positive results incurred when performing multiple hypothesis tests [3–6]. This problem is prominent in the urology literature, as two-thirds of clinical investigations with multiple hypothesis tests fail to account for the increased risk of false positive results [7]. The medical literature contains numerous results of subgroup analyses that were later refuted [1]. The purpose of the present review is to outline criteria, which increase the likelihood that the results of a subgroup analysis are indeed valid, rather than just hypothesis-generating.

☆ The concepts presented were taken in part from the Users' Guide to the Medical Literature and the Evidence-Based Urology in Practice series [1,14].

* Corresponding author. Tel.: +1-713-500-7337; fax: +1-713-500-0546.

E-mail address: steven.canfield@uth.tmc.edu (S.E. Canfield).

1078-1439/$ – see front matter
doi:10.1016/j.urolonc.2011.01.007

2. Clinical scenario

A 60-year-old male presents to your clinic for a second opinion regarding a recent diagnosis of prostate cancer. He recently underwent prostate biopsy for a PSA level of 8.4 ng/ml, and was noted to have Gleason 3+3 = 6 prostate cancer. He is potent and has minimal urinary symptoms. His past medical history is significant only for hypertension managed with a β-blocker, and he has had no prior surgeries. His prostate exam is consistent with clinical T1c disease. He has discussed his cancer management extensively with his family physician and local urologist. His family physician has recommended a course of watchful waiting, whereas his local urologist has recommended that the patient undergo radical prostatectomy. The patient is not interested in radiation therapy. He is seeking a second opinion from a urologic oncologist, and is particularly interested in the evidence for or against treating prostate cancer in men in his age group.

3. Literature search

To find the best evidence regarding the patient's question, you search the medical literature using the PICOT (Patient, Intervention, Comparison, Outcome, and Type of study) method [8,9] derived question,"in men with localized prostate cancer, does radical prostatectomy vs. watchful waiting improve survival?" Since this is a therapy question, a randomized controlled trial would provide the best evidence by minimizing bias in assessing the difference in outcomes [10]. Using PubMed (www.pubmed.gov), you enter the terms "prostate cancer," "radical prostatectomy," and "watchful waiting." Combining the search results for these 3 terms yields 240 articles (search performed October 2010), which is clearly too many to review. You limit the results to "randomized controlled trials," which narrows the list to 36 articles. Scanning these titles, you encounter a manuscript that appears to be a randomized controlled trial of surgery vs. watchful waiting for prostate cancer [11]. You save the article for further review.

4. Critical appraisal

As you review the article, you determine that this indeed is a randomized controlled trial, which appears to meet validity criteria [10,12] and demonstrates a relative risk of death from prostate cancer of 0.56 [95% CI 0.36–0.88] among men undergoing surgery compared with watchful waiting [11]. Interestingly, you note that the authors present the results of a subgroup analysis of men aged less than 65 years, which demonstrated reduced cancer-specific and overall mortality in the surgery group [11]. Aware that the results of subgroup analyses are sometimes problematic, you decide to carefully assess the reported outcomes of men aged less than 65 years in this trial.

5. Assessment of subgroup analysis

When considering the results of a subgroup analysis, we must first convince ourselves that the overall study is valid. Explicit criteria for appraising the results of randomized controlled trials are reviewed in detail elsewhere [12]. For the purposes of this review, we will assume the randomized controlled trial was well conducted, with valid results. Once assured that the overall results are valid, we must consider the characteristics of a subgroup analysis that minimize the chances of a spurious finding (Table 1) [1,13,14]. These characteristics help to maintain the prognostic balance created by the original trial randomization, utilize prespecified hypotheses, use statistical techniques to minimize and measure the probability of chance (Type I error) findings, and consider any detected differences in outcome in the context of the overall evidence.

6. Maintaining prognostic balance

6.1. Is the subgroup variable a characteristic at randomization?

Randomization ideally balances both known and unknown prognostic factors in a clinical trial [1]. A subgroup analysis is more likely to be valid if it is based on factors present at randomization, such as age or gender, rather than factors that may develop during the course of a trial, such as how patients respond to a treatment. In the present case, the authors report 3 subgroup analyses, stratifying by age (less than 65 years of age vs. 65 years of age or older), PSA level at diagnosis (10 ng/ml or lower vs. great than 10 ng/ml), and pre-randomization prostate biopsy Gleason score (less than 7 vs. 7 or greater) [11]. All of these characteristics are present at randomization, and thus have greater likelihood of maintaining prognostic balance within their sub-analysis.

Table 1
Criteria for assessing the validity of a subgroup analysis1 [1,13,14]

Is the subgroup variable a characteristic at randomization? Is the effect suggested by comparisons within rather than between studies?
Was the hypothesis specified a priori?
Was the effect one of a small number of subgroup analyses?
Is the probability small that the observed interaction is due to chance?
Is the interaction consistent across studies? Is the interaction consistent across closely-related outcomes within the study? Is there indirect evidence that supports the hypothesized interaction?