DEFINING BIOCHEMICAL FAILURE FOLLOWING RADIOTHERAPY WITH OR WITHOUT HORMONAL THERAPY IN MEN WITH CLINICALLY LOCALIZED PROSTATE CANCER: RECOMMENDATIONS OF THE RTOG-ASTRO PHOENIX CONSENSUS CONFERENCE

MACK ROACH III, M.D.,* GERALD HANKS, M.D.,† HOWARD THAMES JR., PH.D.,‡ PAUL SCHELLHAMMER, M.D.,§ WILLIAM U. SHIPLEY, || GERALD H. SOKOL, M.D.,¶ AND HOWARD SANDLER, M.D.**

*Departments of Radiation Oncology, University of California San Francisco, San Francisco, CA; †Department of Radiation Oncology, Fox Chase Comprehensive Cancer Center, Philadelphia, PA; ‡Department of Biostatistics and Applied Mathematics, The University of Texas, M.D. Anderson Cancer Center, Houston, TX; §Department of Urology, Eastern Virginia Medical School, Norfolk, VA; ||Department of Radiation Oncology, Harvard Medical School, Massachusetts General Hospital, Boston, MA; ¶U.S. Food and Drug Administration, Rockville, MD; and **Department of Radiation Oncology, University of Michigan, Ann Arbor, MI, USA

In 1996 the American Society for Therapeutic Radiology and Oncology (ASTRO) sponsored a Consensus Conference to establish a definition of biochemical failure after external beam radiotherapy (EBRT). The ASTRO definition defined prostate specific antigen (PSA) failure as occurring after three consecutive PSA rises after a nadir with the date of failure as the point halfway between the nadir date and the first rise or any rise great enough to provoke initiation of therapy. This definition was not linked to clinical progression or survival; it performed poorly in patients undergoing hormonal therapy (HT), and backdating biased the Kaplan-Meier estimates of event-free survival. A second Consensus Conference was sponsored by ASTRO and the Radiation Therapy Oncology Group in Phoenix, Arizona, on January 21, 2005, to revise the ASTRO definition. The panel recommended: (1) a rise by 2 ng/mL or more above the nadir PSA be considered the standard definition for biochemical failure after EBRT with or without HT; (2) the date of failure be determined “at call” (not backdated). They recommended that investigators be allowed to use the ASTRO Consensus Definition after EBRT alone (no hormonal therapy) with strict adherence to guidelines as to “adequate follow-up.” To avoid the artifacts resulting from short follow-up, the reported date of control should be listed as 2 years short of the median follow-up. For example, if the median follow-up is 5 years, control rates at 3 years should be cited. Retaining a strict version of the ASTRO definition would allow comparisons with a large existing body of literature. © 2006 Elsevier Inc.

Biochemical recurrence, Prostate cancer, Radiotherapy, PSA failure.

INTRODUCTION

The serum marker called PSA (prostate specific antigen) is widely used for screening, diagnosing, determining prognosis, and selecting the appropriate treatment for men with clinically localized prostate cancer (1–7). After treatment, PSA is used to determine the effectiveness of treatment. Reports early in the PSA era (early 1990s) tended to emphasize the need for normal values (<4.0 ng/mL) or other threshold values (such as 2 or 1 ng/mL) (8). This lack of standardization made it impossible to compare the results from different institutions.

In 1994 the Board of the American Society for Therapeutic Radiology and Oncology (ASTRO) formed a committee to develop a standard definition for PSA failure after external beam radiotherapy (EBRT). To this end, in 1996 ASTRO sponsored a Consensus Conference to establish a working definition of biochemical failure after EBRT (9). A panel of experts including radiation oncologists, urologists, statisticians, and medical oncologists used the best available evidence at the time and the ASTRO consensus definition was born. This definition provided a standard definition that allowed radiotherapy series from different institutions to be compared.

Stated simply, the ASTRO Consensus Definition (as it came to be called) defined PSA failure as occurring after three consecutive PSA rises after a nadir with the date of failure defined as the point halfway between the nadir date...
and the first rise or any rise great enough to provoke initiation of salvage therapy. The Consensus Panel also went on to say that “it is recommended that series be presented for publication with a minimum period of observation of 24 months” and that “. . . PSA determinations be obtained at 3 to 4 month intervals during the first 2 years after the completion of radiation therapy, and every 6 months thereafter.” Unfortunately, many investigators ignored this last portion of the recommendation. Consequently, there have been many studies published that include patients with inadequate follow-up (as will be discussed in more detail later), leading to inaccurate estimates of long-term outcomes and compromising the robustness of this definition.

Three additional important conclusions were also reached during the first ASTRO Consensus Conference including:

1. “Biochemical failure is not justification per se to initiate additional treatment. It is not equivalent to clinical failure.”
2. “It is however, an appropriate early endpoint for clinical trials.”
3. “No definition off PSA failure has, as yet, been shown to be a surrogate for clinical progression or survival.”

These conclusions reflected the desire for recommendations about therapeutic interventions to be evidence based. They also left open the possibility that “PSA failure” might in some cases be a clinically irrelevant endpoint.

Although creation of the ASTRO Consensus Definition must be viewed as a tremendous success, it became clear with additional data that it was far from an ideal definition. First, backdating seriously biases the Kaplan-Meier estimates of event-free survival and, in a way, that depends on length of follow-up (the bias is worse, the shorter the follow-up) such that reports with different follow-up cannot be compared (10, 11). Second, from the outset it was made explicitly clear that this definition was not linked to clinical progression, survival, or therapeutic interventions. Furthermore, despite the fact that the ASTRO definition was not developed using data from series using hormonal therapy (HT) or brachytherapy (BT), this definition came to be applied in both of these settings as well (12–14). The ASTRO definition of PSA failure also came to be applied to patients treated with nonradiation-based approaches such as radical prostatectomy and cryosurgery (15–17).

To address the shortcomings of the ASTRO Consensus definition, a second Consensus Conference was held on January 2005 in Phoenix, Arizona to formally consider replacing or revising the ASTRO Consensus definition. This conference was jointly sponsored by ASTRO and the Radiation Therapy Oncology Group (RTOG). This report summarizes the data presented, the issues discussed, and the major conclusions reached by the presenters and the panel. It is very important for the readers to note that the definitions proposed are to define success or failure in the context of a population, not an individual. Defining PSA/biochemical success for an individual vs. a population are separate questions with the former being guided by clinical judgment. The definitions chosen for the latter must be such that a computer program can be written to calculate automatically the disease-free status for a large number of patients. They should be used to compare results after EBRT treatment techniques with or without short-term androgen deprivation, but they should not be used for surgically-treated patients, patients undergoing salvage radiotherapy, or patients undergoing cryosurgery.

SUMMARY OF PRESENTATIONS: PSA FAILURE AFTER RADIOTHERAPY: MATERIAL AND METHODS (FORMAT FOR DATA PRESENTED)

The scientific presentations were organized and moderated by the meeting Chair, Dr. Howard Sandler. The program was preceded by a number of phone conferences to clarify the purpose and goals of the meeting, and to select the expert panel and the most appropriate speakers. The speakers were selected because they were recognized experts who had large databases detailing the outcomes after EBRT with sufficient follow-up and patient numbers to justify evidence-based conclusions.

The Panel consisted of senior investigators considered experts on prostate cancer who were charged with questioning the speakers and preparing this report. When possible, we strove for consensus, but compromises were made to reflect the varying opinions.

Introductory presentations

Two introductory presentations preceded the presentation of new data. William Shipley, M.D., reviewed the history of the First Consensus Conference. He reflected on the charge put forward by the ASTRO board to develop a definition for PSA failure in 1994. Dr. Shipley reviewed the methodology used in establishing the ASTRO Consensus definition. He emphasized the fact that an evidence-based approach using pooled multi-institutional data and an expert multidisciplinary panel provided the backbone for this definition (1996). This approach made it distinct from definitions used elsewhere in the literature for radiotherapy and other treatment modalities. He summarized other accomplishments that grew out of the applications of this definition, including the ASTRO-sponsored multi-institutional analysis of PSA failure (1997–1998) using the ASTRO definition (18). He also acknowledged that there were areas in which the ASTRO definition appeared to be inadequate.

Dr. Sandler clarified the need for the current symposium and defined the charge for the Panel. He summarized the major aims as follows: “To seek to standardize criteria of failure that would allow comparisons between reported series and to be useful and relevant to everyday clinical practice.” He reiterated the shortcomings of the ASTRO definition including: (1) being very sensitive to the length of follow-up; (2) it was developed to address EBRT monotherapy; (3) there is a censoring artifact from backdating; (4) there is a potential for false positives secondary to
“benign PSA bounces” associated with HT, BT, and EBRT; and, most importantly, (5) there was a lack of correlation with clinical progression.

Clarification of the charge of the expert panel and sensitivity vs. specificity?

A distinction was recognized between developing a definition of biochemical failure for the purpose of determining whether a patient was cured (yes or no) vs. a definition that correlated with a clinical endpoint (local failure, metastasis, death). It was acknowledged that a definition that substantially improved on the sensitivity and specificity of the ASTRO definition was desirable; however, some panelists believed that both might not be optimally achieved with a single definition. At least one panelist argued strongly for a definition that was directed at prioritizing for an endpoint linked closely to the notion of “cure.” However, such a definition would probably result in more false-positive failures, and a greater number of patients would have to be censored because of inadequate follow-up. Most of the speakers, the experts, and the audience participating favored a definition that could readily be linked to a clinical outcome, and they considered a definition linked to “cure” too rigorous and difficult to precisely define at this time. That is, even “clinically insignificant” recurrences might be counted as failures, reducing the specificity of the definition. This position implies that specificity for a clinical outcome is prioritized. Unfortunately, a definition that emphasizes specificity for clinical outcomes may result in an overestimation of the frequency of cure. An example of such an endpoint would be to define recurrences only when the posttreatment PSA exceeds a very high value, such as 20 ng/mL. With such a definition, the likelihood of detecting bone metastasis on bone scan is much greater than using a definition of >1 ng/mL. Thus, if “cure” is the most important endpoint, sensitivity is prioritized, but if specificity is prioritized then a definition that correlates with local recurrences, regional or distant recurrences, or death from prostate cancer would be prioritized.

### STATISTICAL CONSIDERATIONS

Predictions of disease progression based on the slope of the PSA

Jeremy Taylor, Ph.D., presented data that emphasized the merits of a definition of biochemical failure that was linked to PSA kinetics (19). He presented data that demonstrated that the slope of the PSA when combined with other clinical features (such as the pretreatment PSA, T stage, and radiation dose) could be used to monitor disease progression. Although this approach was seen as elegant, it was perceived as complicated and was based only on patients treated with EBRT, and thus did not address the issue of HT or BT, and has not been evaluated by other institutions. The panel members recognized the logic and merits of slope-based definitions but did not favor this strategy as a replacement for the ASTRO definition (20).

A cooperative group perspective: design and analysis issues of biochemical failure

Michelle L. DeSilvio, Ph.D. (senior statistician, RTOG), discussed issues related the analysis of clinical trials and their design. She discussed the inherent differences between the time of origin in randomized and nonrandomized trials. More important to the issue of defining PSA failure, she presented examples based on the findings of RTOG 9202 and 9413 as to how backdating affected outcomes. Finally, she emphasized that the RTOG used the cumulative incidence method and Gray’s test to assess outcomes and alluded to the development of surrogate endpoints for survival.

Presentation of data for consideration by the panel

Presentation of data were divided into three treatment scenarios: EBRT alone, no androgen deprivation therapy (ADT), EBRT plus ADT, and treatment incorporating permanent prostate seed implants (PPI). Each presentation was limited to 15 min and each scenario was followed by a question-and-answer session led by the panel members.

In search of the perfect failure definition

Dr. Deborah Kuban discussed some of the issues complicating the quest for a perfect definition. Using data pooled from nine prominent institutions, including 4,839 patients with T1–2 disease treated with radiotherapy alone, she reminded the audience of the classic PSA response patterns after radiotherapy. With a median follow-up of 6.3 years, approximately one-third of patients obtained a nadir of <0.5 ng/mL, one-third between 0.5 and 1.0 ng/mL, and the remaining one-third nadirs >1.0 ng/mL. She presented data illustrating how the hazard rates varied by whether the definition incorporated backdating or whether it was based on failure using an “at call” approach. As is shown in Fig. 1a and b, hazard rates appeared to be “front-loaded” (particularly for high-risk patients) when backdating was used. The importance of this point will be revisited in the final recommendation presented below.

Focusing only on patients treated to >70 Gy from this dataset, Dr. Kuban went on to describe how the so-called PSA bounce further complicated matters. Roughly 26% of the patients from this dataset experienced a transient rise in their PSA by 5 years. These transient rises (median level 0.7 ng/mL) occurred at a median follow-up time of 21 months. She provided examples of how the ASTRO definition was sensitive to duration of follow-up because of backdating and showed how this was eliminated by using an “at call” strategy. She also made the point that backdating is not used for any other disease. She highlighted several definitions associated with a higher sensitivity and specificity than the ASTRO definition (11). She also provided data that suggested that the sensitivity and specificity of the so-called current nadir + 2 definition appeared to be the most robust and generalizable for low-, intermediate-, and high-risk patients. In contrast, the ASTRO definition became progressively worse going from low- to high-risk
patients. She also demonstrated that with short follow-up, outcomes using the nadir + 2 definition tended to be more favorable than with the ASTRO definition, but with longer follow-up they tended to be worse (Fig. 2a and b) (21). The importance of this point will also be revisited in the final recommendation presented below.

**Assessment of biochemical endpoints after external beam radiotherapy alone**

Dr. Larry Kestin presented a detailed assessment of various definitions based on 1,457 patients treated for clinical T1–3 to a median dose of 66 Gy at William Beaumont. He also observed that the median posttreatment nadir was 0.7 ng/mL and that at 10 years, 13% had locoregional failure and 14% had distant metastasis. The initiation of salvage HT was not counted as biochemical failure in this analysis. Biochemical failure was considered a prerequisite for defining clinical failure. Using clinical failure as the endpoint, he compared the sensitivity, specificity, and predictive value of 10 different scenarios (Table 1). As is shown in Table 2, patients whose PSA equals or exceeds 3.0 ng/mL at or after achieving an absolute nadir had the
Fig. 2. (a) and (b) were provided courtesy of Dr. Deborah Kuban (The University of Texas M.D. Anderson Cancer Center). Both sets of curves compare the percent disease free of PSA failure at 5 years using the ASTRO Consensus Definition (59%) (thin lines) to those disease free using nadir $+2$ (68%) or nadir $+3$ (72%) (thick lines). Panel c was provided by Dr. Eric Horwitz (Fox Chase Cancer Center) and shows once again that early control rates look better using a definition based on nadir $+2$ ("Houston") but later results favor the ASTRO Consensus definition. Of note, control rates appear to be better earlier with the latter but worse with longer follow-up (personal communication, January 21, 2005, Phoenix, Arizona).
Table 1. Correlation with clinical failure (from Kestin et al.)

<table>
<thead>
<tr>
<th>Definition</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>+ Predictive value (%)</th>
<th>− Predictive value (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 rises (current—resets with repeat)</td>
<td>51</td>
<td>78</td>
<td>31</td>
<td>88</td>
</tr>
<tr>
<td>Absolute nadir ≥1.0 ng/mL</td>
<td>68</td>
<td>66</td>
<td>28</td>
<td>91</td>
</tr>
<tr>
<td>Current nadir + 2 ng/mL</td>
<td>66</td>
<td>77</td>
<td>35</td>
<td>92</td>
</tr>
<tr>
<td>Absolute nadir + 2 ng/mL</td>
<td>64</td>
<td>78</td>
<td>36</td>
<td>92</td>
</tr>
<tr>
<td>Current nadir + 3 ng/mL</td>
<td>58</td>
<td>81</td>
<td>37</td>
<td>91</td>
</tr>
<tr>
<td>Absolute nadir + 3 ng/mL</td>
<td>56</td>
<td>82</td>
<td>38</td>
<td>91</td>
</tr>
<tr>
<td>≥3.0 ng/mL after current nadir</td>
<td>68</td>
<td>73</td>
<td>33</td>
<td>92</td>
</tr>
<tr>
<td>≥3.0 ng/mL after absolute nadir</td>
<td>65</td>
<td>77</td>
<td>35</td>
<td>92</td>
</tr>
<tr>
<td>≥3.0 ng/mL at or after current nadir</td>
<td>74</td>
<td>73</td>
<td>35</td>
<td>94</td>
</tr>
<tr>
<td>≥3.0 ng/mL at or after absolute nadir</td>
<td>72</td>
<td>76</td>
<td>37</td>
<td>93</td>
</tr>
</tbody>
</table>

strongest association with clinical failure at 10 years. On this basis, Dr. Kestin argued for using this definition, suggesting that 44% of patients failing clinically would meet this definition of biochemical failure compared with only 13% biochemically controlled.

External beam monotherapy
Dr. Scott Williams (Peter MacCallum Cancer Centre, Melbourne, Australia) presented data on 1,571 patients with T1–4 disease treated at the Queensland Radium Institute with radiotherapy alone to doses of 52.5 to 68.25 Gy. With a median PSA follow-up of 7.5 years, 71% of the patients from this cohort had experienced biochemical failure using the ASTRO definition and 561 had died. He demonstrated how backdating creates an artificial scenario of the patient being “unable” to die of any cause for the duration of backdating, as well as the duration of follow-up and why he favored the use of a definition based on the nadir + 2 ng/mL.

Biochemical endpoints after external beam radiation
Dr. Eric M. Horwitz presented data from the Fox Chase database. His analysis was based on 1,017 men treated with three-dimensional conformal radiotherapy to a median dose of 74 Gy (range, 61–81 Gy). He made distinctions between defining outcomes for an individual and for a population of patients. He presented data showing that the ASTRO definition overestimates biochemical failure by 20% to 30% (22). He confirmed the observations made by others, that use of the nadir + 2 ng/mL definition results in an apparent improvement in biochemical failure at short follow-up, but a worse outcome long-term follow-up (see Fig. 2c, personal communication, January 21, 2005, Phoenix Arizona).

RADIOTHERAPY COMBINED WITH HORMONAL THERAPY

Biochemical failure and adjuvant androgen deprivation therapy
Dr. Tom Pickles (British Columbia Cancer Agency, Vancouver) discussed the combined use of radiation and ADT including: (1) the impact of testosterone recovery on PSA bounce; (2) when to set “time 0”; and (3) the predictive ability of various PSA failure definitions. He based many of his conclusions on earlier work and an analysis of unpublished data from 1,885 men treated with EBRT (46% with ADT) and 483 men treated with PPI (70% with ADT), with a minimum follow-up of 3 years. He first highlighted the point that 95% of men recover noncastrate testosterone levels at a median of 10 months (23). He also reported that using the nadir + 2 definition improved the accuracy of defining recurrences in patients treated with or without ADT (24). He showed that the ASTRO definition was associated with a greater likelihood of false-positive biochemical failure calls as the result of a PSA bounce when ADT was used compared with a nadir + 2 or 3 (28% vs. <5% for both of the latter). He also concluded that ADT-treated patients should have the duration of testosterone ablation subtracted from their follow-up, in lieu of measuring testosterone levels, 10 months should be used (23).

Dr. Pickles went on to support his conclusions by providing data that directly linked PSA failure with cause-specific survival. He showed that nadir + 2 or 3 ng/mL was better correlated with survival rates than the ASTRO definition. He showed how the ASTRO definition failed to meet proportional hazard assumptions that are considered a basic tenant of multivariate analysis. In contrast, nadir + 2 or 3 met proportional hazard assumptions and concluded that
one of these should be adopted to replace the ASTRO definition.

**PSA failure as a surrogate endpoint in men treated with ADT**

Dr. Richard Valicenti presented data based on RTOG 9202 addressing the role of PSA doubling time (PSADT) as a surrogate for death after treatment with HT and radiation. He noted that the ASTRO Consensus definition failed to meet Prentice’s criteria for surrogacy and investigated PSADT as an alternative. Based on his analysis, PSADT <12 months met Prentice’s criteria and thus should be considered for use in trials as a study endpoint. His presentation (and that of Dr. Taylor’s) provided further validation for the notion that PSA kinetics could bring another level of predictive power to outcome not possible by simply considering a threshold PSA value or small incremental changes (19).

**PSA behavior after prostate BT**

Dr. Anthony Zietman reviewed the issue of biochemical failure after PPI. He referred to a notion he called “the lawnmower principle,” which he described as “the shorter and more evenly cut the grass, the earlier and more clearly you see the weeds.” He was alluding to the fact that PSA values tend to go lower with PPI. He summarized a body of evidence having to do with the magnitude of the PSA values tend to go lower with PPI. He summarized a body of evidence having to do with the magnitude of the PSA values. He noted that the ASTRO Consensus definition failed to meet Prentice’s criteria for surrogacy and investigated PSADT as an alternative. Based on his analysis, PSADT <12 months met Prentice’s criteria and thus should be considered for use in trials as a study endpoint. His presentation (and that of Dr. Taylor’s) provided further validation for the notion that PSA kinetics could bring another level of predictive power to outcome not possible by simply considering a threshold PSA value or small incremental changes (19).

For example, median nadir PSA levels ranged from 0.6 ng/mL for patients treated to 70 Gy to 0.3 for patients treated to 79 Gy to 0.1 ng/mL or less for patients treated with EBRT and PPI.

Dr. Zietman also cited data based on endorectal magnetic resonance spectroscopic imaging that suggest greater degrees of prostate ablation with PPI compared with EBRT (25, 26). Due to the range of PSA responses he argued that one definition of failure might not fit all types of radiation. He summarized a body of literature demonstrating that median magnitude of PSA bounces reported by various series ranged from 0.4 to 0.7 ng/mL (27–29). He also pointed out the fact that the literature tended to suggest that bounces in the PSA were not associated with a more unfavorable outcome. He discussed factors that might be correlated with the likelihood that PSA bounces would occur including the size of the gland, proctitis, pre-implant PSA velocity, and the D90 (27, 30). He concluded that based on a review of available data, the use of a nadir + 2 or 3 definition would probably keep the false-positive PSA failure rate at 3% to 5% or less.

Table 3 summarizes the list of presenters and a key point from each of these speakers. Despite the large number of speakers, consistent themes could be extracted from their presentations. Table 4 summarizes the key criticisms made of the existing ASTRO Consensus. As is shown in this table, backdating was consistently criticized as being uniquely applied to prostate cancer, the

<table>
<thead>
<tr>
<th>Topic, Presenter, Institution:</th>
<th>“Take-home points”</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Radiotherapy alone:</strong></td>
<td></td>
</tr>
<tr>
<td>Deborah Kuban—M.D. Anderson</td>
<td>Reviewed data based on nine institutions (n = 4,839), T1–2, Pts ≥70 Gy, median nadir = 0.6 ng/mL, bounce occurred in 26%. Compared various definitions, presentation favored nadir + 2 ng/mL.</td>
</tr>
<tr>
<td>Larry Kestin—William Beaumont Hospital</td>
<td>Presented data based on 1,457 patients with T1–T3 disease. Detailed analysis of nadir definitions and ROC curves emphasizing clinical failure and merits of using a definition based on an absolute 3 ng/mL.</td>
</tr>
<tr>
<td>Scott Williams—Peter MacCallum, Melbourne, Australia</td>
<td>Presented data on 1,571 patients treated with radiotherapy alone. Argued against backdating and for nadir + 2 ng/mL as reliable and correlated with survival.</td>
</tr>
<tr>
<td>Eric Horwitz—Fox Chase Cancer Center</td>
<td>Presented data on 688 (102 with HT) treated with 3DCRT. Compared ASTRO vs. modified ASTRO vs. nadir + 2 ng/mL. Concluded nadir +2 worked well with and without HT.</td>
</tr>
<tr>
<td>Discussion: 50 min Radiotherapy And Hormonal Therapy:</td>
<td></td>
</tr>
<tr>
<td>Tom Pickles—British Columbia Cancer Agency, Vancouver, Canada</td>
<td>Database from 1,885 patients treated with EBRT (46% + HT) and 483 PPI (70% + HT). Issue of PSA bounces after withdrawal and “false call rate” with HT. ASTRO def. (22%) worst vs. nadir +2 low rate (6%). Nadir +2 vs. ASTRO correlated with 27% vs. 17% surv. Diff. ASTRO non-proportion hazards vs. proportional for nadir +2 or 3.</td>
</tr>
<tr>
<td>Richard Valicenti—Thos. Jefferson</td>
<td>Presented data from RTOG 9202 demonstrating that a PSA doubling time &lt;12 months after PSA failure met Prentice’s criteria for surrogacy for mortality from prostate cancer.</td>
</tr>
<tr>
<td>Discussion: 25 min Brachytherapy:</td>
<td></td>
</tr>
<tr>
<td>Anthony Zietman—Mass Gen Hosp</td>
<td>Discussed the differences in responses between treatment with EBRT and permanent prostate implantation (PPI) 2,325 pts treated before 1999 including time to nadir, frequency and magnitude of PSA bounces (17–43%, 0.7 ng/mL), favored target PSA, 0.7 or nadir + 2 or 3 ng/mL definition.</td>
</tr>
</tbody>
</table>

**Abbreviations:** HT = Hormonal therapy; Mass Gen Hosp = Massachusetts General Hospital.
critical sensitivity to duration of follow-up, and the fact that it violates statistical principles involving proportional hazard estimates. Finally, the ASTRO Consensus definition has been shown to have a lower sensitivity and specificity for clinical outcomes than several alternative definitions (11).

**RECOMMENDATIONS**

We recommend that a rise by 2 ng/mL or more above the nadir PSA (defined as the lowest PSA achieved) be considered as the current standard definition for biochemical failure after radiotherapy with or without short-term hormonal therapy. We recommend that the date of failure be determined “at call” and not backdated. However, confirmed laboratory errors or patients with acute prostatitis effectively treated with antibiotics should not be declared as biochemical failures. Patients not meeting these PSA criteria for failure who undergo salvage therapies (such as ADT, radical prostatectomy or BT, or cryosurgery) should also be declared as failures at the time a positive biopsy is obtained or salvage therapy is administered (whichever comes first). For the purposes of comparing patients treated with radiotherapy and HT to those treated without hormonal therapy, we recommend that this definition be used and that use of the ASTRO consensus definition be considered inappropriate.

We further recommend that investigators be allowed to continue to use the ASTRO Consensus Definition after EBRT or BT alone (no HT) with strict adherence to guidelines as to what is considered adequate follow-up. To avoid the artifacts resulting from short follow-up, we recommend that the stated date of control be listed as 2 years short of the median follow-up (10, 31). For example, if the median follow-up is only 5 years, control rates at 3 years should be cited. Thus, to state a “5-year biochemical control rate” a median follow-up of 7 years would be required. Retaining the ASTRO definition in the context would allow comparisons to be made with a large older body of literature. Adhering to these guidelines would ensure that the results of contemporary series would not be artificially inflated due to short follow-up.

**RATIONALE FOR RECOMMENDATIONS AND PRECAUTIONARY NOTES**

The shortcomings of the ASTRO Consensus definition as noted above includes: (1) the fact that it is very sensitive to the length of follow-up; (2) it was developed to address the issue of EBRT monotherapy; (3) there is a backdating censoring artifact; (4) there is a potential for false positives secondary to “benign PSA bounces” associated particularly with the use of ADT and PPI; (5) it does not meet proportional hazard assumptions (a basic tenant of multivariate analysis); and (6) there was a lack of correlation with clinical progression when developed in 1996.

Despite these problems there are several reasons that the ASTRO Consensus definition should not be completely abandoned. First, there is a huge body of published literature that would immediately become obsolete. Second, prior work has suggested that recurrence rates using the EBRT were comparable with surgery...
when the ASTRO definition was used (32). Replacement of the ASTRO definition with the recommended nadir + 2 definition artificially creates a bias favoring radiotherapy series with short outcomes over surgical outcomes (see Fig. 2a–c). Such comparisons would be inappropriate because the endpoint for surgery is not based on clinical outcome data. Many of the shortcomings associated with using the ASTRO definition can be addressed by strict adherence to rules about adequate follow-up and in patients treated with radiotherapy alone.

Cautionary statement

This is a Consensus Statement and as such represents the opinions of experts based on the best available data in 2006. The recommended definition for biochemical failure reflects an attempt to implement a definition with a greater sensitivity and specificity after EBRT with or without short-term HT. These definitions are not recommended for use in patients treated with other modalities, such as cryosurgery or radical prostatectomy. These definitions (now the so-called “Phoenix Definition” rising from the ashes of ASTRO in Phoenix, Arizona, and the “strict-ASTRO” described above) do not address the issue of how many patients are cured. Longer follow-up and more data are required to address this issue. Physicians should use individualized approaches in managing young patients with a slow-rising PSA who might be a candidate for salvage local therapies (33–37).

REFERENCES

25. Pickett B, Kurnasiewicz J, Pouliot J, et al. Efficacy of high dose external beam radiotherapy (EBRT) compared to perma-
nent prostate implant (PPI) in treating low risk prostate cancer based on endorectal magnetic resonance spectroscopy imaging (MRSI) and PSA. Int J Rad Bio Phys 2004;60(Suppl): S185–S186.


