INSTITUTE FOR CLINICAL AND ECONOMIC REVIEW

FINAL APPRAISAL DOCUMENT

ACTIVE SURVEILLANCE & RADICAL PROSTATECTOMY FOR THE MANAGEMENT OF LOW-RISK, CLINICALLY-LOCALIZED PROSTATE CANCER

September 11, 2009

Daniel A. Ollendorf, MPH, ARM	Chief Review Officer
Julia Hayes, MD	Lead Decision Scientist
Pamela McMahon, PhD	Senior Decision Scientist
Steven D. Pearson, MD, MSc	President, ICER

EXECUTIVE SUMMARY

Introduction

Prostate cancer is the second leading cause of cancer deaths and the seventh overall cause of death in men in the United States (Centers for Disease Control and Prevention, 2008). Given that most new cases are diagnosed at an early, localized stage, significant attention has been focused on understanding the risks and benefits of alternative management strategies for patients with low-risk disease. The major options include active surveillance and various forms of radiation therapy and surgery. Most men in the United States choose one of the definitive forms of treatment, but data to compare the long-term risks and benefits of active surveillance and each of the definitive treatment options are limited, placing great emphasis on the need for objective sources of guidance to help clinicians and patients engage in active shared decision-making.

ICER has previously appraised the comparative clinical effectiveness and value of 4 forms of radiation therapy: intensity-modulated radiation therapy (IMRT), low-dose-rate interstitial brachytherapy, proton beam therapy (PBT), and three-dimensional conformal radiation therapy (3D-CRT). This appraisal focuses on active surveillance as well as the major approaches to radical prostatectomy—namely, the traditional "open" approach, minimally invasive laparoscopic prostatectomy, and robot-assisted laparoscopic prostatectomy. It must be emphasized that this review is relevant only for considerations of the management of localized, low-risk disease; the evidence and clinical tradeoffs involved in the treatment of intermediate- or high-risk prostate cancer would differ substantially.

For active surveillance and radical prostatectomy there are several key questions that have served to frame this review:

1) The impact of active surveillance and radical prostatectomy on survival, freedom from disease progression/recurrence, and quality of life relative to alternative management options
2) The relative rates of symptom progression and treatment-related complications and side effects
3) The effects of variation in practice and surgeon experience, frequently called the "learning curve," on patient outcomes for the different versions of radical prostatectomy
4) The patient clinical characteristics and individual values that may influence the relative risks and benefits of these alternative management options
5) The cost-effectiveness and budget impact of active surveillance and radical prostatectomy relative to alternative management options

Because these management options represent very different pathways of care, with potentially important differences in short and long-term risks and benefits, all health care decision makers will benefit from a formal appraisal of the comparative clinical effectiveness and comparative value of active surveillance and the alternative surgical prostatectomy options for clinically-localized, low-risk prostate cancer.

Alternative Treatment Options

Active Surveillance

Because of the limited aggressiveness of many localized prostate cancers, active surveillance is viewed by most experts as a reasonable strategy for many men with low-risk prostate cancer (NCCN, 2008). Nonetheless, in the U.S. active surveillance is infrequently used; data from 1990-2006 in the CaPSURE registry suggest that active surveillance is employed in <10% of low-risk patients (Cooperberg, 2007).

The term 'watchful waiting' is sometimes used interchangeably with active surveillance. However, watchful waiting was first coined during an era when many men were first diagnosed with prostate cancer not through PSA screening but through presentation with obstructive urinary symptoms or a palpable nodule. It has been estimated that PSA screening detects prostate cancers an average of 9 years before clinical diagnosis in the absence of screening, and therefore patients with PSA-screen-detected disease will have a much more favorable outcome, even without treatment, than patients diagnosed clinically in earlier watchful waiting studies (Parker, 2004).

Following the publication of randomized controlled trials that showed a 10-year survival advantage for radical prostatectomy over this earlier form of watchful waiting (Bill-Axelson 2005, 2008), current practice has shifted away from a relatively passive watchful waiting approach towards what is a much more active program of surveillance via repeated PSA tests and prostate biopsies, with definitive treatment triggered by any sign of biochemical or pathological progression. The major differences between the older version of watchful waiting and the modern approach to active surveillance are illustrated in the graphic below, based on a prototypical set of criteria used in the UK (Parker, 2004).

Contrasts between active surveillance and watchful waiting.

	Active Surveillance	***Watchful Waiting***
Primary Aim	*To individualize treatment*	*To avoid treatment*
Patient Characteristics	*Fit for radical treatment; age 50-80*	*Age >70 or life expectancy <15 years*
Tumor Characteristics	*T1-T2, Gleason ≤7, Initial PSA <15*	*Any T stage, Gleason ≤7, Any PSA*
Monitoring	*Frequent PSA testing, Repeat biopsies*	*PSA testing unimportant, No repeat biopsies*
Indications for Treatment	*Short PSA doubling time, Upgrading on biopsy*	*Symptomatic progression*
Treatment Timing	*Early*	*Delayed*
Treatment Intent	*Curative*	*Palliative*

Source: Parker C. Active surveillance: towards a new paradigm in the management of early prostate cancer. *Lancet Oncol* 2004;5:101-6.

Professional guidelines have identified multiple criteria that define candidacy for AS; a common definition is based on a Gleason score (a measure of tumor aggression) of 6 or less, PSA levels 10 ng/ml or less, and a stage between T1c and T2a (NCCN, 2009). Patients with Gleason scores of 7 are also often considered eligible for active surveillance. Other criteria that may be used include 33% or fewer positive cores (biopsy samples), or 50% or fewer single-core involvement. When a patient opts for active surveillance, he is put on a regular monitoring schedule. While there is no universal standard protocol for active surveillance, monitoring schedules often include serial PSA blood tests every 3-6 months, digital rectal exams (DRE) every 3-6 months, and a repeat biopsy at one year followed by subsequent biopsies every 3-5 years thereafter (Klotz, 2008). Other monitoring tests that have been employed include bone scans and CT scans of the abdomen and pelvis to monitor for metastases, as well as transrectal ultrasounds in combination with DRE to assess for progression of local disease or urinary symptoms (Choo, 2002).

Thresholds to trigger definitive treatment in patients on active surveillance are also not universally agreed upon. A rapid rate of PSA increase, or the "PSA velocity", is used by some physicians as an indicator of aggressive disease. Others consider the doubling of a PSA level within 3-4 years (i.e., "PSA doubling time") to be the most reliable indicator of disease progression. Still others contend that results of repeat biopsies provide the best predictor of disease progression. Because the natural history of prostate cancer is poorly understood, clinicians often consider all of these potential triggers to judge when to advise patients that definitive treatment should be initiated.

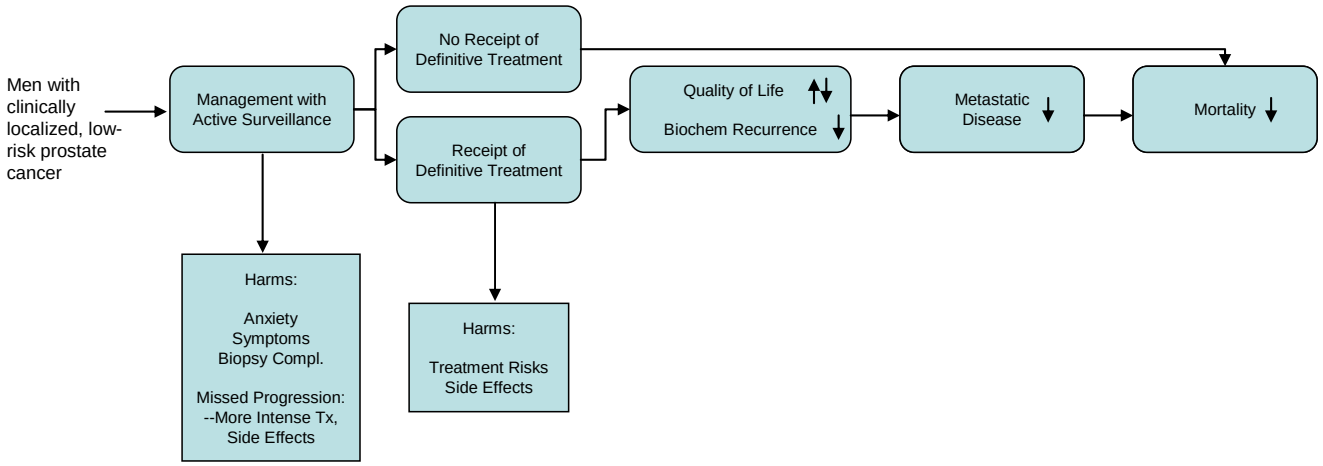
Radical Prostatectomy

Radical prostatectomy has long been an option for the treatment of prostate cancer. The procedure involves the surgical removal of the prostate gland, seminal vesicles, and, in some cases, lymph nodes under general anesthesia; an inpatient hospital stay of 1-4 days' duration is typical. Radical prostatectomy is usually performed when the cancer is localized to the prostate. Candidates for surgery are generally in good overall health with a life expectancy of at least 10 years. There are 3 major surgical approaches employed in radical prostatectomy: radical retropubic prostatectomy (i.e., the traditional "open" surgical approach), as well as two minimally-invasive surgical approaches, laparoscopic radical prostatectomy and robot-assisted laparoscopic prostatectomy. Modern applications of both open and minimally-invasive prostatectomy also involve the use of "nerve-sparing" techniques in an attempt to preserve post-surgical erectile function.

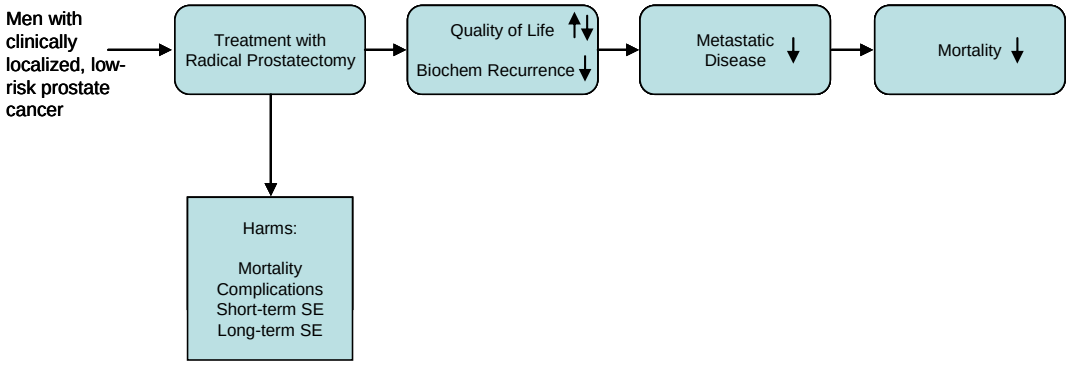
Utilization of laparoscopic and, in particular, of robot-assisted procedures have increased dramatically in recent years. Between 2003 and 2005, utilization of minimally-invasive techniques among Medicare beneficiaries grew from 12.2% to 31.4% (Hu, JCO, 2008), a change likely to have been driven primarily by growth in robot-assisted surgery (Blute, 2008). Advocates for these techniques cite potentially reduced blood loss as well as shorter hospital stays and recovery time as advantages over open prostatectomy (Berryhill, 2008). There is a steep learning curve associated with these procedures, however, as surgeons must adjust to reduced range of motion, discontinuity between real and visible movement, and reduced tactile feedback (Rassweiler, 2006).

Analytic Framework for Evaluation of Active Surveillance and Radical Prostatectomy
The analytic framework for this review is shown in two figures on the following page; one each for active surveillance and radical prostatectomy. There are little to no data directly demonstrating the impact of these therapies on overall patient survival, so judgments about the effectiveness of these interventions must rest almost exclusively upon consideration of the strength of surrogate endpoints as well as evaluation of treatment-associated risks. Note that the figures below are intended to convey the conceptual links involved in evaluating outcomes of these management options, and are not intended to depict a clinical pathway that all patients would transit through. A separate depiction of clinical pathways is available in the Economic Model section of this report (see Section 8).

Analytic Framework: Active Surveillance in Prostate Cancer Treatment



Analytic Framework: Radical Prostatectomy in Prostate Cancer Treatment



The link between biochemical evidence of disease recurrence and survival has been the subject of much debate. Because of the slow growth of most prostate cancers, and the consequent need for extremely long follow-up periods to measure survival accurately, biochemical recurrence, or "failure," as marked by changes in PSA levels following treatment, is widely used as a predictor of survival. Some evidence suggests that biochemical failure is an appropriate surrogate in certain subgroups, such as high-risk patients younger than 75 years (Kwan, 2003). Questions remain, however, regarding biochemical failure's prognostic ability for other patients. Nonetheless, biochemical failure has gained broad consensus among clinicians and researchers as a valid surrogate outcome. Clinicians use it as a trigger for decisions to employ adjuvant or salvage therapy following prostatectomy, and its role as a surrogate measure in research will endure due to the practical barriers to conducting large-scale trials of sufficient duration to measure disease-specific and overall mortality.

Evidence on Comparative Clinical Effectiveness

Data Quality

A total of 111 studies met all entry criteria for review. Randomized controlled trials do not exist that compare measures of benefit and/or harm between active surveillance and radical prostatectomy. Randomized evidence is limited to the Scandinavian randomized controlled trial of radical prostatectomy vs. watchful waiting (Bill-Axelson, 2005) as well as a single-center study comparing open and laparoscopic prostatectomy (Guazzoni, 2006). The remaining studies included retrospective cohort studies of cancer registries or claims databases as well as case series predominantly from single academic sites. While some of the surgical case series included comparisons to historical or contemporaneous controls receiving an alternative surgical approach, comparisons between groups are problematic for multiple reasons, including selection bias, changes in surgical protocols over time, differential follow-up for newer vs. older approaches, and changes in measurement of prostate cancer severity.

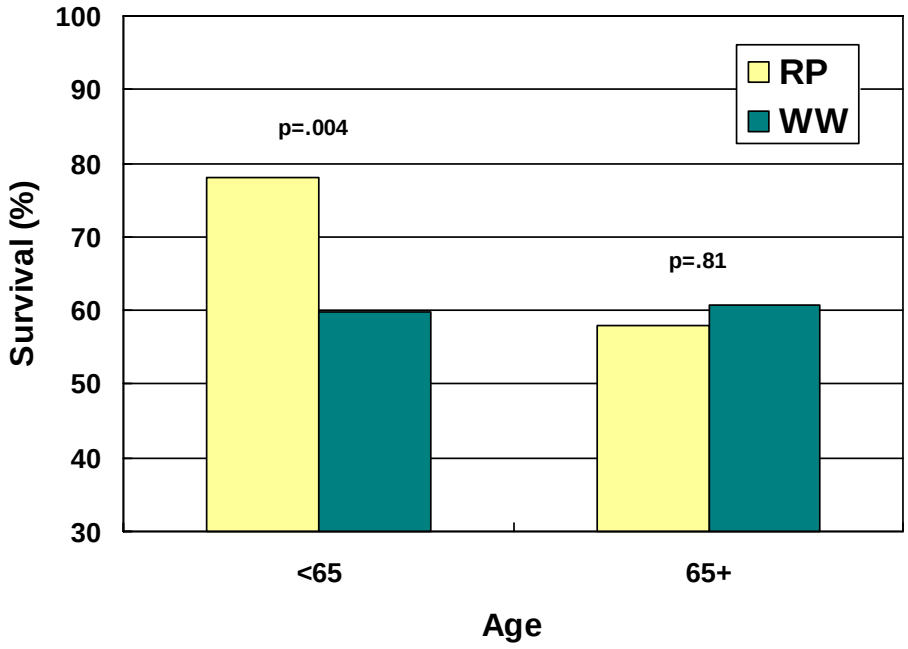
Data on active surveillance are also limited, given its relatively recent evolution from watchful waiting. The longest reported median follow-up is 7 years (vs. 20-30 years in some watchful waiting studies); in addition, only one active surveillance study involved a comparison to a treatment alternative, a contemporaneous comparison to a watchful waiting cohort (Hardie, 2005). The lack of a substantive body of data on active surveillance outcomes beyond 5-7 years limits the level of certainty that can be achieved in comparisons of clinical effectiveness, particularly for younger patients (<65 years old) who would be expected to live an additional 20 years or more.

Survival and Freedom from Biochemical Failure

There are no data available to directly compare the impact of active surveillance vs. radical prostatectomy on overall survival. Some articles draw inferences of a lower boundary for active surveillance from older randomized controlled data on watchful waiting vs.

prostatectomy, in which the results indicated a survival benefit for surgery in men under age 65, but not in those 65 and older (see Figure ES1 below).

Figure ES1. 12-year overall survival by age and treatment arm, SPCG-4 trial.
Source: Bill-Axelson, JNCI, 2008

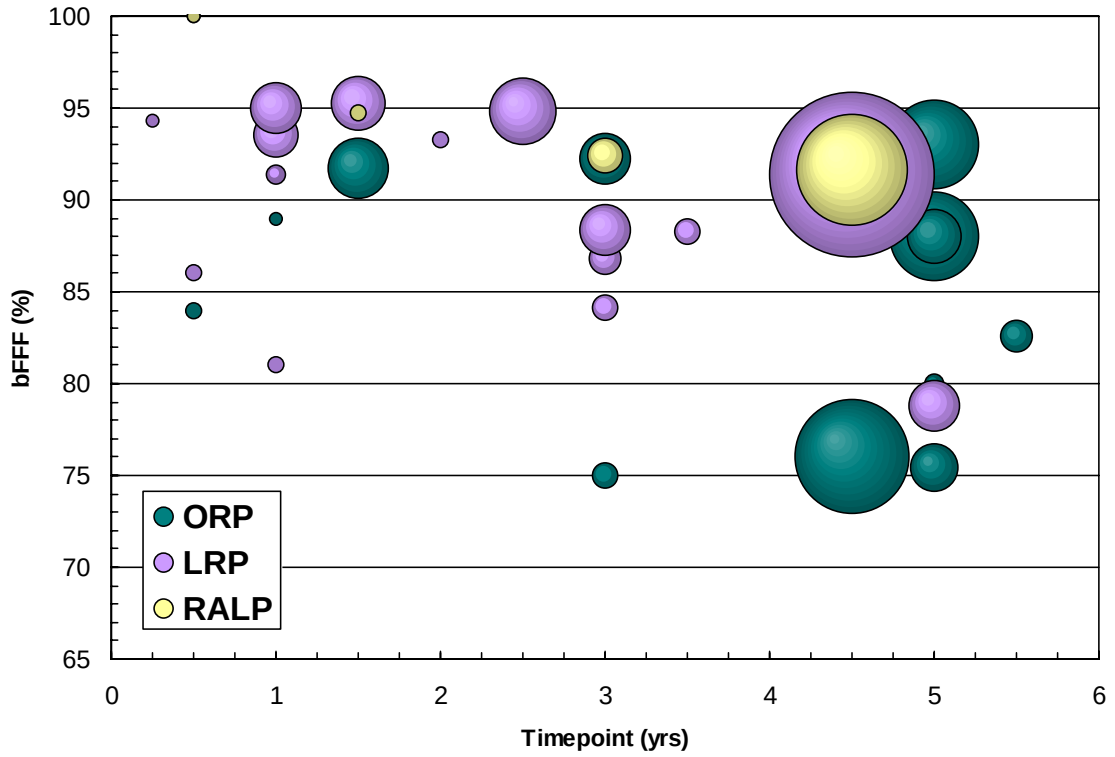


While there are no studies that directly compare active surveillance to radical prostatectomy, 5-year survival rates in published case series are comparable (range: 84-99%). No studies comparing the impact of different surgical approaches on overall survival have been published.

Similar evidence limitations characterize findings on disease-specific survival. No studies have directly compared active surveillance to radical prostatectomy, nor have any evaluated the impact of surgical approach on this outcome. However, published case series estimates of five-year disease-specific survival for both active surveillance and radical prostatectomy largely overlap in a range from 86-100%.

Comparisons of freedom from biochemical failure (bFFF) following prostatectomy is complicated by the use of variable definitions of biochemical failure, as well as differences in duration of follow-up, pathological tumor staging, and other patient and/or study characteristics. When studies were limited to those with sufficient follow-up to report 3-year or longer-term estimates, findings were similar across surgical approaches (see Figure ES2 on the following page).

Figure ES2. Biochemical freedom from failure, by surgical approach and timepoint.



ORP: Open prostatectomy; LRP: Laparoscopic prostatectomy; RALP: Robot-assisted laparoscopic prostatectomy; NOTE: Bubble size used to illustrate study sample size

Treatment-Free Survival in Active Surveillance

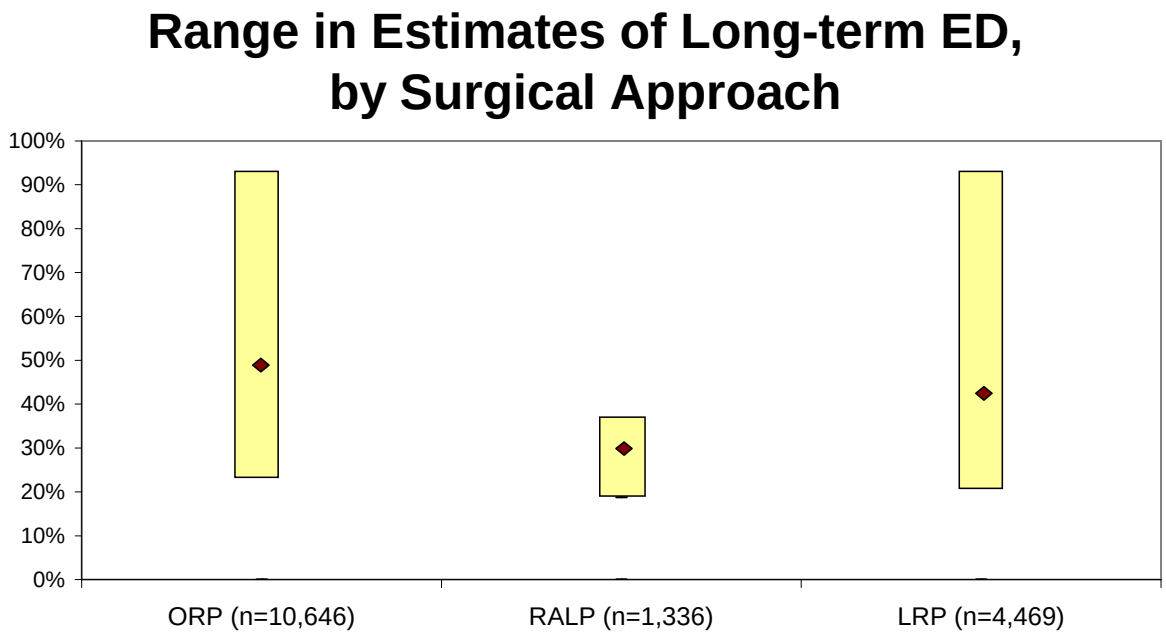
Approximately 25%-50% of patients who begin active surveillance will ultimately receive some form of treatment within 5-10 years. Very limited data suggest that approximately one-third to one-half of decisions to initiate definitive treatment are due to patient choice and not because of clinical or pathologic progression. Sparse data show that Gleason grade progression occurs in 5%-40% of men over time, with nearly all grade change from 3+3 at diagnosis to 3+4 disease after re-biopsy (Dall'Era, 2008; Carter, 2007; Klotz, 2007). In addition, between 25%-65% of men are found to have a completely benign pathology on first re-biopsy (Soloway, 2008). The clinical significance of Gleason grade progression or regression on surveillance biopsies is unknown (Dall'Era, 2009). Because active surveillance differs fundamentally from watchful waiting in its inclusion of the possibility of treatment with curative intent, the proportion of patients ultimately receiving treatment cannot be directly compared across these two approaches.

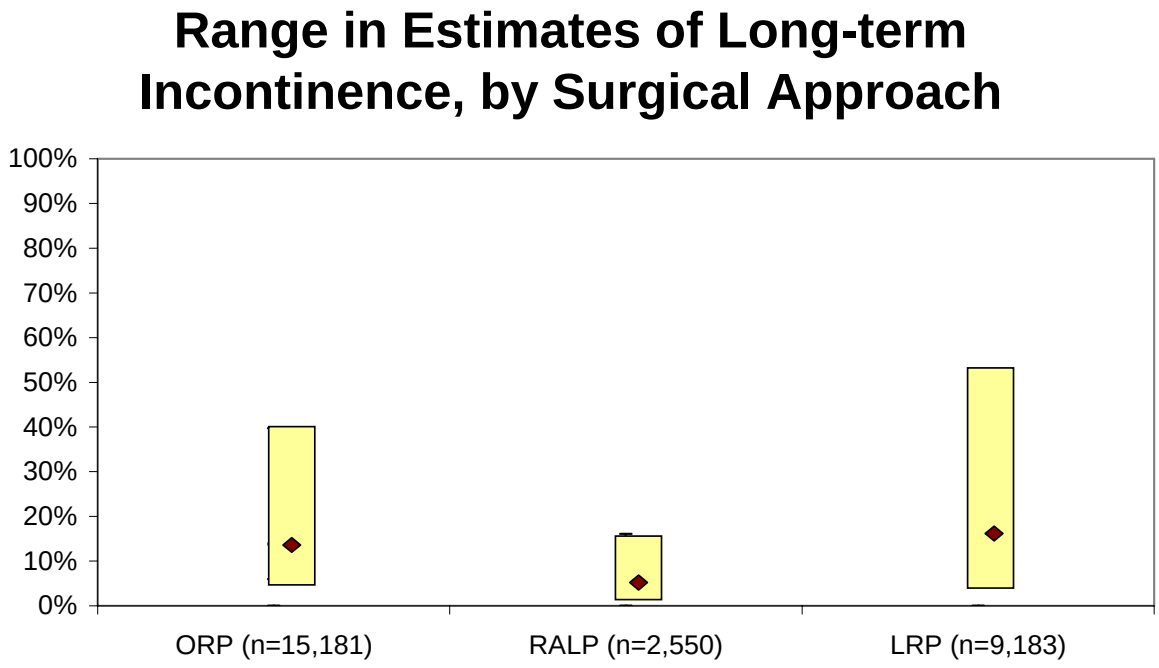
Potential Harms: Radical Prostatectomy

Nearly all information on potential harms of radical prostatectomy comes from individual case series, necessitating indirect comparisons across surgical practices and patient populations that differ in demographic and clinical characteristics, study timeframe, measurement of outcome, and other characteristics. Because of these concerns, the ICER

review assigns no degree of certainty to nominal differences in the published rates of harms by surgical approach. As examples, the variability in estimates for long-term erectile dysfunction and incontinence are illustrated below.

Pooled data on these harms are presented for informational purposes alone in Table ES1 on page 14.





NOTE: Diamonds represent pooled mean rate; rectangles represent full range of estimates

Peri-Operative Complications

Intra- or peri-operative mortality is rare across all surgical approaches to prostatectomy, with a risk of approximately 0.4% for 65 year-old men. While rates differ somewhat by patient age, the risk is well below 1% in all age groups, and does not differ materially by surgical approach.

Data on complications are extremely variable due to differences in measures, patient populations, surgeon experience, and other factors. A rough estimation based on pooled data suggest that the risk of major complications, including DVT/PE, MI, and stroke, is approximately 3-4% and does not appear to materially differ across surgical approaches. The risk of minor peri-operative complications such as UTI or wound infection is approximately 8-9%. The limited comparative data available suggest that minimally-invasive prostatectomy performed by experienced surgeons may be associated with lower rates of minor peri-operative complications, but interpretation of these data is complicated by the younger age of patients undergoing minimally-invasive techniques, and complication rates appear significantly higher among surgeons with limited experience with the newer techniques. Operative blood loss is lower in minimally-invasive approaches, as are associated transfusion requirements, but there is no evidence of a reduced risk of major hemorrhage.

Urethral Stricture, Incontinence, and Erectile Dysfunction

The risk of urethral stricture varies considerably in the published literature, with estimates ranging from less than 1% to 15%. Some evidence suggests that the risk of stricture has declined significantly over time, as all surgical techniques have evolved. Evidence is conflicting on the impact of minimally-invasive surgery on stricture rates; studies of employer and Medicare claims data have indicated reduced risk of stricture from minimally-invasive prostatectomy among younger patients, while for unclear reasons an increased risk was observed in older men (Hu, JCO, 2008; Hu, J Urol, 2008).

Incontinence remains a significant concern among patients undergoing radical prostatectomy, regardless of surgical approach. Approximately 40% of patients will have incontinence at 3 months post-surgery. This side effect appears to resolve in many patients 12 or more months after surgery, but our analysis suggests that between 10-15% of men will still require occasional or consistent pad use at 12-24 months. Evaluation of differences by surgical approach is problematic for many reasons, including differential follow-up and patient age; however, existing data do not suggest a substantial difference in the risk for acute or chronic incontinence by surgical approach.

Both short- and long-term erectile dysfunction (ED) are also significant concerns among men undergoing radical prostatectomy. Approximately 70% of men experience ED in the first three months following surgery. ED improves over the course of the year, but at 12 months following surgery approximately 35% of men who were potent prior to bilateral nerve-sparing surgery will still have ED. Rates of ED among men receiving unilateral or non nerve-sparing surgery are between 50-80%. Data are not sufficient to perform reliable comparisons of ED across different surgical approaches.

Table ES1. Reported harms of radical prostatectomy, by surgical approach.

Measure	ORP	LRP	RALP
Peri-Operative			
*Mortality**	Studies: 62 Pooled: 0.4% Range: 0.0-0.7%	Studies: 62 Pooled: 0.4% Range: 0.0-0.7%	Studies: 62 Pooled: 0.4% Range: 0.0-0.7%
Major Comp	Studies: 20 Pooled†: 4.7% (3.7%, 5.7%) Range: 2.1%-28.6%	Studies: 21 Pooled: 3.5% (2.4%, 4.6%) Range: 0.0%-36.6%	Studies: 12 Pooled: 2.5% (1.4%, 3.6%) Range: 0.0%-7.8%
Minor Comp	Studies: 20 Pooled: 9.5% (3.3%, 15.7%) Range: 0.3%-25.3%	Studies: 21 Pooled: 7.8% (6.1%, 9.4%) Range: 0.0%-23.5%	Studies: 12 Pooled: 5.3% (3.1%, 7.4%) Range: 0.5%-15.0%
Conversion	N/A	Studies: 22 Pooled: 0.4% (-0.1%, 0.9%) Range: 0.0%-3.7%	Studies: 14 Pooled: 0.1% (-0.1%, 0.3%) Range: 0.0%-2.3%
+ *Margins* *(pT2)* *(pT3)*	Studies: 14 Pooled: 16.8% (13.2%, 20.4%) Range: 6.0%-34.2% Pooled: 45.2% (35.5%, 55.0%) Range: 9.1%-84.6%	Studies: 25 Pooled: 13.9% (12.1%, 15.7%) Range: 4.7%-30.2%) Pooled: 39.3% (35.0%, 43.5%) Range: 16.7%-71.0%	Studies: 10 Pooled: 10.5% (8.1%, 12.8%) Range: 2.5%-20.0% Pooled: 35.4% (26.6%, 44.2%) Range: 13.0%-66.7%
Side Effects			
Urethral Stricture	Studies: 13 Pooled: 3.4% (2.5%, 4.4%) Range: 0.4%-19.8%	Studies: 16 Pooled: 0.3% (0.1%, 0.6%) Range: 0.0%-6.4%	Studies: 7 Pooled: 1.3% (0.3%, 2.4%) Range: 0.0%-2.3%
Urinary Incontinence	*Acute* Studies: 7 Pooled: 46.7% (25.1%, 68.2%) Range: 25.0%-90.2% *Long-term* Studies: 17 Pooled: 12.7% (9.6%, 15.8%) Range: 6.1%-39.5%	*Acute* Studies: 11 Pooled: 43.0% (23.9%, 62.0%) Range: 8.0%-89.8% *Long-term* Studies: 19 Pooled: 17.3% (13.7%, 20.8%) Range: 5.0%-52.2%	*Acute* Studies: 7 Pooled: 28.9% (13.6%, 44.2%) Range: 6.7%-65.2% *Long-term* Studies: 7 Pooled: 7.3% (2.9%, 11.7%) Range: 2.9%-16.0%
Erectile Dysfunction	*Acute* Studies: 5 Pooled: 76.8% (66.2%, 87.4%) Range: 62.5%-95.1% *Long-term* Studies: 16 Pooled: 45.3% (38.7%, 51.9%) Range: 24.0%-90.0%	*Acute* Studies: 10 Pooled: 71.4% (60.2%, 82.6%) Range: 57.7%-94.7% *Long-term* Studies: 17 Pooled: 41.4% (34.6%, 48.3%) Range: 21.9%-91.2%	*Acute* Studies: 3 Pooled: 59.1% (43.2%, 74.9%) Range: 46.9%-71.7% *Long-term* Studies: 7 Pooled: 26.3% (22.2%, 30.4%) Range: 18.8%-35.0%

*Meta-analysis of mortality data by surgical approach infeasible due to large number of zero values
NOTES: ORP: Open radical prostatectomy; LRP: Laparoscopic radical prostatectomy; RALP: Robot-assisted radical prostatectomy
†From random-effects meta-analysis (with 95% confidence intervals)

Potential Harms: Active Surveillance

Biopsy-related Complications

Data are extremely limited on the incidence and severity of complications arising from initial or repeat prostate biopsy during active surveillance. In addition, measurement of the type and severity of complications varies greatly by study. Nevertheless, prostate biopsy appears to be a relatively safe procedure. The majority of complications reported are transient and self-limiting, such as pain, rectal bleeding, hematuria, and hematospermia.

Data from the largest of these studies, an examination of initial and repeat biopsy in over 1,000 men enrolled in a prospective study of prostate cancer detection (Djavan, 2001), indicated that the incidence of the two most serious complications requiring intervention, namely urosepsis and acute urinary retention, was 0.1% and 2.6% respectively.

Patient Anxiety

While the possibility exists that obstructive urinary symptoms and erectile dysfunction may worsen during active surveillance, data are available only from the Toronto cohort, where findings suggested a rate of symptomatic progression of approximately 3% at a median of 3.75 years of follow-up (Choo, 2004). Limited data on symptom progression are available from watchful waiting studies, but the evidence is not comparable due to the older age and advanced cancer characteristics of these cohorts.

Uncertainty regarding cancer progression while on active surveillance does have the potential to impact patient anxiety. While anxiety levels do appear to predict receipt of definitive treatment among men on surveillance programs, limited data from the active surveillance and watchful waiting literature suggest that overall anxiety levels do not differ between men who have selected these regimens and those who choose initial definitive treatment with radiation therapy or surgery.

Learning Curve: Radical Prostatectomy

There is a substantial learning curve for all forms of radical prostatectomy; cases performed by inexperienced surgeons tend to have higher rates of complications, side effects, disease recurrence, and need for subsequent treatment.

The impact of the learning curve can be observed across multiple measures of surgical outcomes. For example, the average rate of conversion from minimally-invasive to open prostatectomy due to failure of the minimally-invasive approach is less than 1%; however, rates as high as 14% have been observed among surgeons who are relatively inexperienced with the technique. Similarly, evidence from claims-based studies suggest that rates of salvage radiation or hormonal therapy after prostatectomy, treatments often indicative of positive surgical margins, are over 2 times greater among surgeons with a low volume of minimally-invasive surgeries vs. high-volume surgeons (Hu, JCO, 2008).

Given the strength of the data linking surgeon experience to broad ranges of complications and side effects, variability between surgeons and institutions is likely a more important

predictor of patient outcomes than any difference that might be due to the surgical approach selected. For example, if the ranges of side effects found in the ICER systematic review are assumed to arise solely from differences in surgical expertise, a surgeon performing at the 75th percentile among his or her peers would have a combined major complication rate of approximately 2-3%, with long-term rates of ED at 30-35% and incontinence at 5-7%. These complication and side effect rates would be significantly lower than those of surgeons operating at the 25th percentile, whose patients would suffer major complications at 10-12%, ED at 50-60%, and incontinence at 15-20%. Not all of the variation in published outcomes can be ascribed to surgical expertise, but the data do suggest that variation in surgical performance is a critical feature in any evaluation of the comparative effectiveness of radical prostatectomy to active surveillance or other interventions for localized prostate cancer.

Hospital Costs and Efficiency: Open vs. Robotic Prostatectomy

In the U.S., Medicare reimbursement for all 3 surgical approaches to prostatectomy is similar, with the only difference being a $500 higher payment for the CPT code associated with minimally-invasive approaches. However, costs to the hospital differ substantially, as acquisition, maintenance, and supply costs for laparoscopic guidance and robot systems add significantly to the costs of providing these services. For example, recent estimates of the cost of a robotic surgical system include acquisition costs of $1.6 million, annual maintenance costs of $100,000-$200,000 and disposables costs of $2,000-$3,000 per case (Lotan, 2004; Joseph, 2008; Quang, 2007). Minimally-invasive prostatectomy has been associated in the literature with reductions in the length of hospital stay of 2-3 days compared to open prostatectomy, but the use of clinical pathways in many institutions has also resulted in shortened length of stay and reduced transfusion requirements to levels that are indistinguishable by surgical approach (Farnham, 2006; Nelson, 2007). Published evidence indicates that operating-room time is longer with minimally-invasive surgery; findings from our systematic review indicated average operative time of approximately 3 hours for open prostatectomy, vs. 4-4.5 hours for minimally-invasive techniques. This is due to the technical complexity of minimally-invasive procedures; as with other operative outcomes, there is some evidence that operative times shorten as surgeons gain more experience with minimally-invasive techniques (Ficarra, 2009).

Analysis of Comparative Value

We used findings from our systematic review on clinical effectiveness to inform a primary cost-utility analysis of active surveillance and radical prostatectomy in 65-year-old men with localized prostate cancer. Due to the emphasis many clinicians place on age and life expectancy at the time of diagnosis, we also performed an analysis with a cohort of 55-year-old men, as well as multiple sensitivity analyses examining potential variations in relative differences in outcomes and costs between the various treatment strategies. Utilities (i.e., the value, between 0 and 1, placed on quality of life in a particular state of health) for patients with individual side effects or side-effect combinations were obtained from published literature. Costs of surveillance, surgery, complications, and side effects were based on national Medicare payment rates for relevant services; the costs of patient time

associated with these services were also estimated using national wage rates. Two alternative analyses were performed using actual third-party payer costs obtained from private health plans in the United States.

The "base case" economic model developed for this analysis was framed with the assumption that active surveillance and radical prostatectomy achieve comparable overall mortality rates in men with low-risk, localized prostate cancer. This assumption was based on the existing data on active surveillance which, through 5-7 years of follow-up, does not suggest any decrement in overall or cancer-specific survival for active surveillance compared to prostatectomy. However, because the existing data cannot exclude some chance of a survival benefit for prostatectomy in later years, an alternative scenario was created in which the prostate cancer-specific mortality of active surveillance patients is set at 2.5% higher than radical prostatectomy at 10 years (and persisting for the remainder of the patient's life). This survival advantage for prostatectomy reflects another assumption: that any possible survival advantage for prostatectomy over active surveillance will be, at most, approximately half of the absolute survival difference seen in earlier trials of watchful waiting, when patients were largely diagnosed clinically, as opposed to through PSA testing, and when the protocol did not involve close surveillance with the goal of initiating curative treatment for early biochemical or histological signs of progression (Bill-Axelson, 2005).

In the model, patients aged 65 or older starting on active surveillance who experience progression to intermediate-risk disease are assumed to receive intensity-modulated radiation therapy (IMRT) with short-term androgen deprivation therapy (ADT); patients over 65 who opt for definitive treatment for reasons other than grade progression receive IMRT alone. Radical prostatectomy was assumed as the definitive treatment of choice for all active surveillance patients if they are under age 65 at the time definitive treatment is begun.

Other key assumptions within the economic model are shown below in Table ES2 on the following page and are discussed more fully in the body of this review.

Table ES2. Major assumptions of the ICER economic model

ASSUMPTION	RATIONALE & SOURCE
• No men will die of prostate cancer within 6 months of diagnosis	Low prostate cancer specific mortality in low-risk patients -ICER Review
• All men who recur after treatment recur biochemically	Patients monitored closely by PSA after treatment -ICER Review
• Progression from recurrence to metastatic disease to death identical regardless of treatment	No proven disease-related benefit to one treatment over another -ICER Review
• Men on AS who receive treatment have = risk of CaP death as men treated initially	No studies with sufficient follow up to suggest mortality benefit or harm to AS -ICER Review
• Treatment after AS is RP if <65 or IMRT (w/ or w/o ADT) if ≥65	Mortality benefit to RP vs. WW limited to men <65 yo -Bill-Axelson, 2005
• No men treated with RP receive adjuvant/salvage XRT	<10% low-risk CaP have positive margins at RP -Louie-Johnsun, 2009; Griffin, 2007 Use of salvage XRT in men with low-risk disease <15% -Lu-Yao, 1996; Grossfeld, 1998

NOTES: AS: Active surveillance; RP: Radical prostatectomy; IMRT: Intensity-modulated radiation therapy; XRT: external beam radiation therapy; WW: Watchful waiting; CaP: Prostate cancer

Base Case Model Results

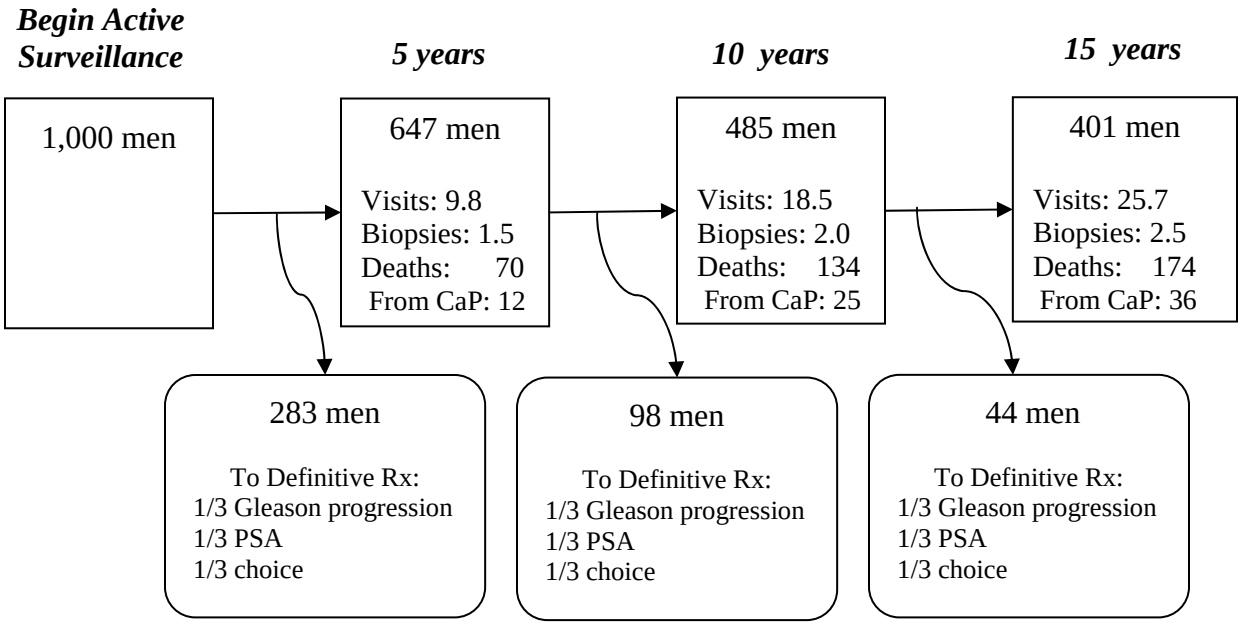
Clinical Outcomes

Under the assumption that active surveillance and prostatectomy confer equal survival, men at age 65 with low-risk prostate cancer have an additional life expectancy of approximately 16 years with either form of management. Complications and side effects reduce the final total of quality-adjusted life years.

A flowchart based on model results of the progression of visits, biopsies, and decisions to enter into definitive treatment for patients aged 65 beginning on active surveillance is displayed in Figure ES3 below. Among men on active surveillance, the likelihood of receiving definitive treatment is 28%, 45%, and 54% after 5, 10, and 15 years respectively, and 61% over a lifetime. Decisions to opt for definitive treatment are driven by approximately equal proportions of men with Gleason progression on surveillance biopsy,

increases in PSA doubling time or other PSA-related findings, and patient choice without objective findings of disease progression. By year 15, men on active surveillance will have had, on average, approximately 26 visits and 2.5 biopsies. These numbers reflect an average that includes the experience of the entire cohort; after adjustment for attrition due to mortality, more than 50% of patients originally on surveillance will have moved into definitive treatment by 15 years.

Figure ES3. Schematic flowchart of 5, 10- , and 15-year cumulative visits, biopsies, all-cause and disease-specific mortality, and treatment decisions of among a cohort of 65 year-old men beginning active surveillance for low-risk, clinically-localized prostate cancer. Data derived from ICER decision-analytic model.



For men treated with radical prostatectomy, the model results showed a risk of peri-operative death of 0.4%, reflecting the parameter input from the ICER systematic review. The risk of developing new ED following radical prostatectomy is 31%; the risk of urinary incontinence is 9%. These estimates are lower than those produced by the ICER review, as they reflect incidence over and above the underlying risk of these conditions due to age and comorbidity. Inclusion of higher estimates would likely magnify the quality-of-life effects already observed with active surveillance (see below). Among the men on active surveillance who ultimately receive IMRT, there are small increased risks for ED, incontinence, and proctitis. A table summarizing the key rates for both short-term and long-term side effects for radical prostatectomy and active surveillance in men aged 65 and 55 is shown in Table ES3 on the following page.

Table ES3. Comparative Value Evidence Table (CVET): Lifetime clinical outcomes for 65- and 55-year-old men with clinically-localized, low-risk prostate cancer.

Outcome (%, except where noted)	Active Surveillance	Open Radical Prostatectomy	Difference (ORP-AS)
Age 65 Years			
Prog. to treatment	61.1%	100.0%	38.9%
Peri-operative death	N/A	0.4%	N/A
Minor complications	0.2%	9.5%	9.3%
Major complications	0.0%	4.8%	4.8%
Treatment-related SE			
Incontinence	3.6%	8.6%	5.0%
ED	5.3%	30.7%	25.4%
GI (from IMRT)	2.7%	N/A	N/A
Prostate cancer death	9.0%	9.0%	0.0%
Life years (mean)	16.0	16.0	0.0
QALYs (mean)	8.97	7.82	(1.2)
Age 55 Years			
Prog. to treatment*	72.1%	100.0%	28.0%
Peri-operative death	0.1%	0.4%	0.3%
Minor complications	2.8%	9.5%	6.7%
Major complications	1.1%	4.8%	3.7%
Treatment-related SE			
Incontinence	6.5%	10.4%	4.0%
ED	13.7%	35.7%	22.0%
GI (from IMRT)	2.0%	N/A	N/A
Prostate cancer death	16.0%	16.0%	0.0%
Life years (mean)	22.0	22.0	0.0
QALYs (mean)	11.54	10.33	(1.2)

NOTES: SE: side effects; ED: erectile dysfunction; GI: gastrointestinal; IMRT: intensity-modulated radiation therapy; QALYs: quality-adjusted life years
*In this younger-age population, 30% of treated patients receive radical prostatectomy

Costs

The initial cost of treatment with radical prostatectomy was $13,553, a figure that represents a Medicare payment rate based on the estimated proportion of cases that are uncomplicated (86%), and that are associated with minor (9.5%) or major (4.8%) complications. Active surveillance is less expensive than radical prostatectomy in the early years following diagnosis, but the results of pathway cost analyses provided by the model suggest that over a lifetime the average costs for active surveillance in 65-year-old men are estimated to be approximately $2,000 higher than for radical prostatectomy ($30,422 vs. $28,348). A breakdown of costs for each pathway is shown in Table ES4 on the following page. As can be seen, the similarity in total lifetime costs between active surveillance and radical prostatectomy is largely driven by the costs of definitive treatment with IMRT or radical prostatectomy ultimately received by over 60% of patients who start active surveillance. As

shown in alternative analyses in the body of the review, active surveillance becomes less costly overall compared to radical prostatectomy if less-expensive brachytherapy is used for definitive treatment in lieu of IMRT.

Table ES4. Comparative Value Evidence Table (CVET): Average lifetime costs for 65- and 55-year-old men with clinically-localized, low-risk prostate cancer.

Cost ($)	Active Surveillance	Open Radical Prostatectomy	Difference (ORP-AS)
Age 65 Years			
Year 1 treatment	4,228	13,553	9,325
Services	4,809	4,624	(185)
Visits	3,382	4,624	1,241
Biopsies	1,427	N/A	N/A
Definitive Rx (IMRT)	14,327	N/A	N/A
Patient time	8,156	6,150	(2,006)
Short-term SE	270	1,477	1,207
Long-term SE	589	786	196
TOTAL			
Undiscounted	38,542	33,589	(4,953)
Discounted	30,422	28,348	(2,074)
Age 55 Years			
Year 1 treatment	3,796	14,496	10,700
Services	5,530	5,213	(317)
Visits	3,848	5,213	1,365
Biopsies	1,682	N/A	N/A
Definitive Rx (IMRT/RP)	13,986	N/A	N/A
Patient time	12,226	9,132	(3,094)
Short-term SE	647	1,468	821
Long-term SE	545	718	173
TOTAL			
Undiscounted	46,690	40,699	(5,991)
Discounted	33,642	31,440	(2,202)

NOTES: SE: side effects; IMRT: intensity-modulated radiation therapy; RP: radical prostatectomy
Component costs presented for illustrative purposes, and will not sum to discounted total

Incremental Cost-effectiveness

As shown on the following page in Table ES5 the model results demonstrated that the avoidance or delay of surgery-related harms afforded by active surveillance translates into a substantial net benefit in quality of life, as this strategy produces an additional 1.15 quality-adjusted years of life compared to immediate radical prostatectomy. Findings were similar for 55-year-old men as well, and for simplicity only the results for 65-year-old men are shown in Table ES5.

Active surveillance was thus found to have higher clinical effectiveness, as measured in quality-adjusted life years, than radical prostatectomy, at an additional lifetime cost of $2,074. The formal incremental cost-effectiveness ratio of active surveillance is $1,803 per QALY gained. For 55 year-old men, active surveillance remained substantially more effective, and cost differences were similar (incremental cost-effectiveness ratio: $1,820 per QALY gained).

Table ES5. Lifetime quality-adjusted life expectancy and costs for 65-year-old men with clinically-localized, low-risk prostate cancer, by treatment type.

Strategy	QALYs	Incremental QALYs	Cost	Incremental Cost	Cost/QALY
Open RP	7.82	Reference	$28,348	reference	
AS	8.97	1.15	$30,422	$2,074	$1,803

NOTES: RP: radical prostatectomy; AS: active surveillance; QALY: quality-adjusted life years

The findings noted above include estimates for reductions in quality-of-life from having undergone radical prostatectomy or remaining on active surveillance, even if no complications or side effects are encountered. When quality-of-life reductions were limited to those arising from side effects, complications, and symptoms only, QALY differences were not as pronounced (10.75 vs. 10.09 for active surveillance and radical prostatectomy respectively), resulting in approximately 8 additional months of quality-adjusted life expectancy for patients in active surveillance at an incremental cost-effectiveness ratio of $3,142 per QALY gained.

Under the alternative model framework in which there is an assumed absolute prostate cancer-specific mortality difference of 2.5% at 10 years in favor of radical prostatectomy, the model results indicated that active surveillance still produced substantially more QALYs on a population basis than prostatectomy, with an additional 0.99 QALYs per patient.

Uncertainty in the base-case model results was assessed through a probabilistic sensitivity analysis. Average costs and QALYs were determined from 100,000 individual-level runs of the model with a unique set of draws from distributions around costs, utilities, and probabilities. Average QALYs for radical prostatectomy were lower than the *lowest* estimates for active surveillance in approximately 30% of the runs; QALYs were higher than the *highest* estimates in another 16%. Formal results of the probabilistic sensitivity analysis are described and discussed in the body of this review.

Alternative Scenarios and Sensitivity Analyses

The body of the report includes the results of several other alternative scenarios, along with the results of numerous one-way sensitivity analyses. Among the key findings was that if the definitive treatment received by patients beginning active surveillance is changed from IMRT to brachytherapy, the active surveillance pathway retains its higher QALY

production but becomes approximately $4,000 less expensive than radical prostatectomy. In all alternative scenarios and sensitivity analyses, active surveillance generated higher QALYs than radical prostatectomy. And, even in scenarios in which costs were increased for active surveillance, such as when representative private payer costs were examined, the absolute lifetime cost difference between active surveillance and prostatectomy remained small, leading to cost savings for active surveillance or incremental cost-effectiveness ratios for active surveillance well below $10,000 per QALY.

Open radical prostatectomy vs. robot-assisted laparoscopic prostatectomy
The findings of the systematic review, and assumptions about costs in the economic model, meant that our base case analysis was not constructed to compare different surgical approaches for radical prostatectomy. We did perform an alternative analysis assuming "maximal" effectiveness for robotic vs. open prostatectomy--in other words, if all nominal differences of the pooled results in the systematic review were considered true differences. Using these estimates, an 8-week gain in QALYs would be realized for robot-assisted surgery from reduced rates of complications and side effects. In addition, lifetime cost savings of approximately $1,700 would be obtained with robotic prostatectomy. It is important to note that the cost estimates used in this analysis are based on Medicare payments for these surgical techniques, and do not take into account the substantial differences in acquisition cost, maintenance, and supplies between the surgical approaches.

Findings on Economic Impact
A summary of the economic impact of active surveillance and radical prostatectomy can be found in Table ES6 on the following page; for the purposes of simplicity, results are presented only for 65 year-old men. Along with the incremental cost-effectiveness ratio, the Table provides evidence on estimated budget impact for a cohort of 1,000 prostate cancer patients over a two-year period. In the first two-year period following diagnosis, a strategy of active surveillance would save nearly $8 million dollars under current Medicare reimbursement rates; a savings of over $13 million dollars would be expected under one of the private payer actual cost scenarios evaluated.

Table ES6 on the following page also presents a hypothetical "fixed budget tradeoff" suggesting potential annual incremental health system spending for doctors and nurses that could be afforded with the potential cost savings achievable by shifting care for 1,000 patients from radical prostatectomy to active surveillance. These figures ignore the downstream costs of definitive treatment for many patients started on active surveillance, and are presented primarily in the spirit of exploring different frameworks through which evidence on value can be presented to decision-makers.

Table ES6. Comparative Value Evidence Table (CVET): Additional findings on value for 65-year-old men with clinically-localized, low-risk prostate cancer.

Measure	Active Surveillance	Open Radical Prostatectomy	Difference (ORP-AS)
1. Service Impact			
Visits	35.9	37.2	1.3
Biopsies	2.8	0.0	(2.8)
Pathway Total	38.8	37.2	(1.6)
2. Cost per Life-Year Saved	N/A		(equivalent survival)
3. Cost per QALY Gained	$1,803		
SA 1: 55 yo men	$1,820		
SA 2: Private-pay estimate A	$3,434		
4. Budget Impact (per 1,000, 2 years)	$5,809,000	$13,591,000	$7,782,000
Using Private-Pay Estimate A	$8,721,000	$22,028,000	$13,307,000
5. Fixed Budget Tradeoffs (Annual)		38.9	*Nurse FTEs @ $100K each*
		19.5	*MD FTEs @ $200K each*
		2.4	*Robotic Surgical System @ $1.6M each*

NOTES: QALY: Quality-adjusted life year; FTE: Full-time equivalent

ICER Evidence Review Group Deliberation

The ICER Evidence Review Group deliberation (see section starting on page 32 for membership and details) focused on many important issues regarding the evidence provided by the ICER review. Major points of discussion are shown in the numbered points below.

1) *Lack of comparative data on overall and disease-specific survival for active surveillance and radical prostatectomy is not a reason to assume NO mortality differences.* While it was recognized that candidate populations for active surveillance and watchful waiting differ in many important respects, including level of risk at diagnosis and intensity of monitoring, several clinicians suggested that the survival data currently available for AS are too premature to compare this approach to definitive treatment. The lack of longer-term data was viewed as particularly relevant for the level of certainty regarding the clinical effectiveness of active surveillance for patients under age 65. Others, however, felt confident that active surveillance's effectiveness would be no worse, and likely better, than watchful waiting, which appears to produce survival equivalent to surgery in older patients. Nevertheless, to explore this uncertainty, ICER created an alternative scenario for the model that assumes a disease-specific survival benefit for radical prostatectomy equivalent to approximately one-half that observed in the SPCG-4 trial (i.e., absolute difference of 2.5% at 10 years and persisting for life).

2) *Variability in surgical practice should receive greater emphasis in the report.* The report has been revised to further highlight (a) the lack of training and competency standards for

newer surgical approaches; (b) variability in outcomes by surgeon and institution; and (c) the impact of the learning curve on all potential surgical outcomes. An additional sensitivity analysis has also been conducted to explore the impact of complication and side-effect rates at the 25th vs. 75th percentile of the observed distribution.

3) *The review and economic model should include the harms of repeat biopsy among patients on active surveillance.* Consistent with the approach taken for complications and side effects of surgery, focus was placed on those outcomes that necessitated significant intervention—namely, cases of urosepsis and acute urinary retention. The costs and utility decrements associated with these complications have been added to the model.

4) *The assumption that active surveillance maintains a constant monitoring intensity over time is incorrect.* Most clinicians in the group felt that physician visits would certainly decrease in frequency as patients remain on surveillance. The model has been adjusted to decrease visit frequency from quarterly to semi-annually after one year.

5) *Any tabular display of nominal differences in outcomes by surgical approach carries the risk of these differences being perceived as "real", despite the presence of cautionary language in the text.* Presentation of these findings has been modified to include graphic displays of selected outcomes to illustrate the wide range and significant overlap observed. The presentation of nominal reported differences has been preserved; this does not imply superiority of any one technique over another, but was done to meet the stated goal of exploring all relevant outcomes of each surgical approach independently.

6) *The disutility estimates for health states defined by the absence of complications or side effects were called into question.* Many on the ERG felt that the original estimates of quality-of-life impact from simply having undergone radical prostatectomy or remaining on surveillance were too severe and imbalanced between management options. Utility estimates have been revised to use the same methodologic approach (i.e., time tradeoff) for both radical prostatectomy and active surveillance; in addition, alternative analyses have been conducted in which reductions in quality-of-life are estimated for side effects, complications, and symptoms alone.

7) *It was noted that presented data on differences in hospital length of stay and operative time between surgical approaches may be based on historical data for open surgery, and that improvements in clinical pathways have minimized differences between surgical approaches.* Further discussion of this issue has been added to the report.

8) *The current analysis has not considered the effects of undergrading of cancer severity and risk on prostate biopsy.*
The draft report has previously noted the phenomenon of Gleason progression on re-biopsy, with reported rates of 30-40% over time; to date, however, the long-term effects of such undergrading are unknown. In addition, several other studies have observed rates of *benign* pathology of 25-60% at re-biopsy, with future consequences that are again unknown. It is possible the improvements in biopsy techniques may reduce the

occurrence of these phenomena. In any event, sensitivity analyses were conducted in this appraisal with an assumed difference in cancer-specific mortality, with findings that did not materially differ from those of primary analyses.

Discussion of ICER Integrated Evidence Ratings

Background on the ICER rating methodology, including descriptions of the rating categories for comparative clinical effectiveness and comparative value, can be found on page 35 of this Executive Summary.

The discussions of the assignment of ICER ratings for comparative clinical effectiveness and for comparative value were conducted separately for comparisons of active surveillance to open radical prostatectomy and for robotic vs. open prostatectomy respectively. Surgery's status as the "reference" category for these ratings in no way implies that ICER considers it a more proven technology or the standard of care. Rather, the rating system is designed to make two-way comparisons, and it is standard practice to make the most frequently-employed or longest-standing therapy the "reference" intervention.

Seven of 15 participants felt that the evidence was sufficient to rate active surveillance's clinical effectiveness as at least "Comparable" to open prostatectomy, while 4 participants felt that the evidence base, while promising, was still too thin to label active surveillance at a level higher than "Unproven with Potential". Of the remaining participants, 2 were undecided between "U" and "C", while 2 others considered the evidence "Insufficient" to make a determination. With regard to comparative value, 6/13 and 4/13 participants rated active surveillance as definitively "High" and "Reasonable/Comparable" respectively, while the remainder were undecided between these two levels (note: Insufficient ratings do not carry a value designation).

Responses were more varied when robotic vs. open prostatectomy was the comparison of interest. Four of 15 participants rated the evidence on clinical effectiveness for robotic surgery to be "Insufficient", 3 rated the procedure as "Unproven with Potential", and 2 each rated the procedure as "Comparable" and "Incremental" respectively. The remainder of participants rated robotic surgery along the continuum between "Insufficient" and "Incremental". Based on a payer perspective that included patient time costs, the majority of participants rated robotic surgery's comparative value as "Reasonable/Comparable", although one participant allowed for a possible "High" designation, and another allowed for the possibility of a "Low" value designation.

The input of the ERG is advisory to ICER; the ultimate rating is made after independent discussion and reflection on the entirety of the review as well as associated meetings. The final ICER ratings are shown on the following pages. Further description of ICER's rationale for the ratings is provided after the figures.

ICER Integrated Evidence Rating™: Active Surveillance vs. Open Radical Prostatectomy for Clinically-Localized, Low-Risk Prostate Cancer

The Comparative Clinical Effectiveness of active surveillance among patients with clinically-localized, low-risk prostate cancer is rated as:

- ***C --- Comparable for patients aged 65 years*, based on similar survival in comparative studies of watchful waiting and radical prostatectomy and early data from active surveillance series showing similar or better outcomes in comparison to watchful waiting**
- ***U --- Unproven with Potential for patients aged 55 years*, based on differences in survival from the watchful waiting/radical prostatectomy studies and early data from active surveillance series showing similar outcomes to radical prostatectomy series**

The Comparative Value of active surveillance among patients with clinically-localized, low-risk prostate cancer is rated as:

- **a --- High**

The Integrated Evidence Ratings:

Ca for patients aged 65 years

Ua for patients aged 55 years

Comparative Clinical Effectiveness	a High	b Reasonable/Comp	c Low
Superior: A	Aa	Ab	Ac
Incremental: B	Ba	Bb	Bc
Comparable: C	AS=Ca (65 y)	Cb	Cc
Inferior: D	Da	Db	Dc
Unproven/Potential: U/P	AS=Ua (55 y)	Ub	Uc
Insufficient: I	I	I	I

Comparative Value

ICER Integrated Evidence Rating™: Robot-Assisted vs. Open Radical Prostatectomy for Clinically-Localized, Low-Risk Prostate Cancer

The Comparative Clinical Effectiveness of robot-assisted radical prostatectomy among patients with clinically-localized, low-risk prostate cancer is rated as:

- **U --- Unproven with Potential**

The Comparative Value of active surveillance among patients with clinically-localized, low-risk prostate cancer is rated as:

- **b --- Reasonable/Comparable***

The Integrated Evidence Rating = Ub

*** Based on current 3rd party reimbursement policy that does not materially distinguish between surgical approaches**

Comparative Clinical Effectiveness	a	b	c
Superior: A	Aa	Ab	Ac
Incremental: B	Ba	Bb	Bc
Comparable: C	Ca	Cb	Cc
Inferior: D	Da	Db	Dc
Unproven/Potential: U/P	Ua	RALP=Ub	Uc
Insufficient: I	I	I	I
	High	Reasonable/Comp	Low
		Comparative Value	

Description of Rationale for ICER Integrated Evidence Ratings

ICER opted to create two ratings in comparing active surveillance and radical prostatectomy: one for "younger" patients (aged 55), and one for "older" patients (aged 65). These are very rough categories meant to capture and reflect the different level of certainty ICER felt the evidence could support for different age cohorts given the relatively short-term data on active surveillance. The rating for patients aged 65 reflects a high level of certainty that the net health benefit of active surveillance is comparable to that provided by radical prostatectomy, as well as the possibility that active surveillance may in fact provide an incremental benefit once more mature data become available. The data from the randomized trial of watchful waiting did not show any significant difference in overall or prostate cancer-specific mortality for men over age 65, and the 5-7 year data available on active surveillance, combined with the "earlier" identification of prostate cancer through PSA testing, creates a persuasive argument that the comparative clinical effectiveness of active surveillance for older, low-risk, localized prostate cancer patients is very comparable to that of radical prostatectomy. In fact, these data would likely have resulted in a "comparable" rating even if the comparison was between *watchful waiting* and radical prostatectomy. Although the model suggested higher average QALYs for active surveillance, which might support a judgment of "incremental" comparative clinical effectiveness, ICER judged that the relative variation in many factors, including surgical expertise and individual patient utilities for various side effects, made "comparable" the most reasonable designation for comparative clinical effectiveness.

The rating for patients aged 55 reflects the lower, "moderate" certainty that ICER judged the evidence supported for a designation of a comparable or incrementally better net health benefit for active surveillance. The ICER rating reflects our judgment that, even though the data are limited, there is reasonable certainty that modern protocols for active surveillance produce mortality outcomes not substantially inferior to radical prostatectomy, while maintaining the quality-of-life advantages of having many patients never require definitive treatment.

The comparative value rating for active surveillance vs. radical prostatectomy reflects consideration of the model results showing low incremental cost-effectiveness ratios, significant near-term cost savings for patients opting for active surveillance, and the fact that under several alternative reimbursement and treatment scenarios, active surveillance appears to be both more effective and cost-saving. In particular, input from the ERG made it clear that many patients begun on active surveillance, even if aged 65 or older, would be treated with prostatectomy or brachytherapy instead of IMRT should they desire or require definitive treatment. The selection of less expensive definitive treatment is a key variable in the modeling of active surveillance, and one that ICER felt supported an overall judgment of a comparative value rating of "high value."

The ratings for the comparison of open and robot-assisted radical prostatectomy are based on the consideration that even though the data on outcomes of patients treated with the robot-assisted technique are extremely limited, the technique is a variation on radical

prostatectomy and not an entirely new modality of treatment; accordingly, ICER felt there was "moderate" certainty that the comparative clinical effectiveness of robot-assisted prostatectomy is at least comparable, and perhaps "incremental" to the traditional open procedure. Given that third-party payment for robot-assisted prostatectomy is currently set at essentially the same rate as that for open radical prostatectomy, it seemed most logical to rate the comparative value "reasonable/comparable." It is possible that the high acquisition cost and the increased marginal costs of robot-assisted surgery will be factored into reimbursements in the future; there is also the countervailing argument that, at least in some institutions, robot-assisted prostatectomy can aid progress toward a lower length of hospital stay. How these various costs play out for different stakeholders in the health care system is difficult to estimate, reinforcing our judgment that a suitable designation for comparative value at this time is "reasonable/comparable."

Sample Physician-Patient Script

Discussing the evidence on potential risks and benefits of active surveillance and immediate treatment options is a central element of shared decision-making between patients, clinicians, and families. ICER offers the script below as an example of how clinicians could initiate a conversation with a 65 year-old male with prostate cancer that would foster consideration of the findings of this evidence review. Conveying this amount of information in one conversation may not be practicable or appropriate for many patients; nor is the text intended to be prescriptive regarding any one management option. The intent is to foster discussion, and to suggest only one of many styles through which clinicians can empower their patients to share in the consideration of the evidence on reasonable clinical alternatives and to help them choose the option that will reflect their broader best interests.

"I know you've looked at your options, including surgery, radiation treatment, or what is called "active surveillance" for your prostate cancer. We've talked a little bit about these options already. Today let's go further. First, I'd like you to know that evidence reviews and national expert groups have concluded that – for men like you with low-risk prostate cancer – there is no evidence that any of these approaches is any better than the others at keeping you healthy and extending your life. Active surveillance is, therefore, a very reasonable option for you to consider, as it would allow you to delay, for many years in some circumstances, the discomfort and side effects that may occur with treatment. On the other, hand, some men do opt for treatment right away, so let's talk about the radiation options and surgery. We have had more years of experience with brachytherapy; IMRT has been in use for about 8 years; and proton beam therapy is fairly new for prostate cancer so we have far less data on its longer-term outcomes. Similarly, there is more evidence on traditional "open" prostatectomy, and less on laparoscopic and robotic approaches. These options have potential advantages and disadvantages with regard to possible side effects of treatment, which you should take some time thinking about. In addition, all of these options require differing amounts of procedure and monitoring time as well as numbers of visits to the doctor. And, some are more expensive than others, both for your own out-of-pocket costs and for your health plan. Before we run through some of these pros and cons together, let me stop here to see if you have any questions or if you've heard anything about any of these options that you'd like to discuss…."

Evidence Review Group Members
The Evidence Review Group (ERG) is an independent group brought together by ICER and composed of academic experts, patients, clinicians, epidemiologists, ethicists, and medical policy representatives of stakeholder groups including health plans and manufacturers.

The purpose of the ERG is to guide and help interpret the entire appraisal process. Members of the ERG are first convened to function as a "scoping committee" for the appraisal. During this phase the key questions for the appraisal are outlined, including elements such as the appropriate comparator technologies, patient outcomes of interest, patient subpopulations for which clinical and cost-effectiveness may vary systematically, time horizon for outcomes, and key aspects of the existing data that must be taken into account during the appraisal. The ERG may be divided into sub-committees that advise the ICER appraisal team at the mid-point of the appraisal on the early findings and challenges encountered. All of the ERG members listed below participated in scoping and/or mid-cycle activities, but not all were able to participate in the final ERG meeting.

At the final ERG meeting, members are asked to declare any interests in the technology or its comparator(s), or other potential influences on their expertise (listed below). The ERG meeting allows for in-depth deliberation on the findings of the ICER appraisal document and provides an opportunity for comment on the determination of the ICER integrated evidence rating. Although the ERG helps guide the final determination of the ICER Integrated Evidence Rating™, the final rating is ultimately a judgment made by ICER, and individual members of the ERG should not be viewed in any way as having endorsed this appraisal.

ERG Participant Name and Affiliation	Potential Influences on Expertise
Peter Albertsen, MD, MS Professor of Surgery, Chief and Program Director, Division of Urology University of Connecticut Health Center Director, Connecticut Institute for Clinical and Transitional Science	Member of Blue Cross/Blue Shield Technology Assessment Panel; member of steering committee for ProtecT trial (active surveillance vs. immediate surgery or radiation)
John Ayanian, MD, MPP Professor of Medicine & Health Care Policy Harvard Medical School & Brigham & Women's Hospital Professor of Health Policy & Management Harvard School of Public Health	None
Peter Carroll, MD Professor and Chair, Department of Urology Ken and Donna Derr-Chevron Distinguished Professor University of California, San Francisco	None

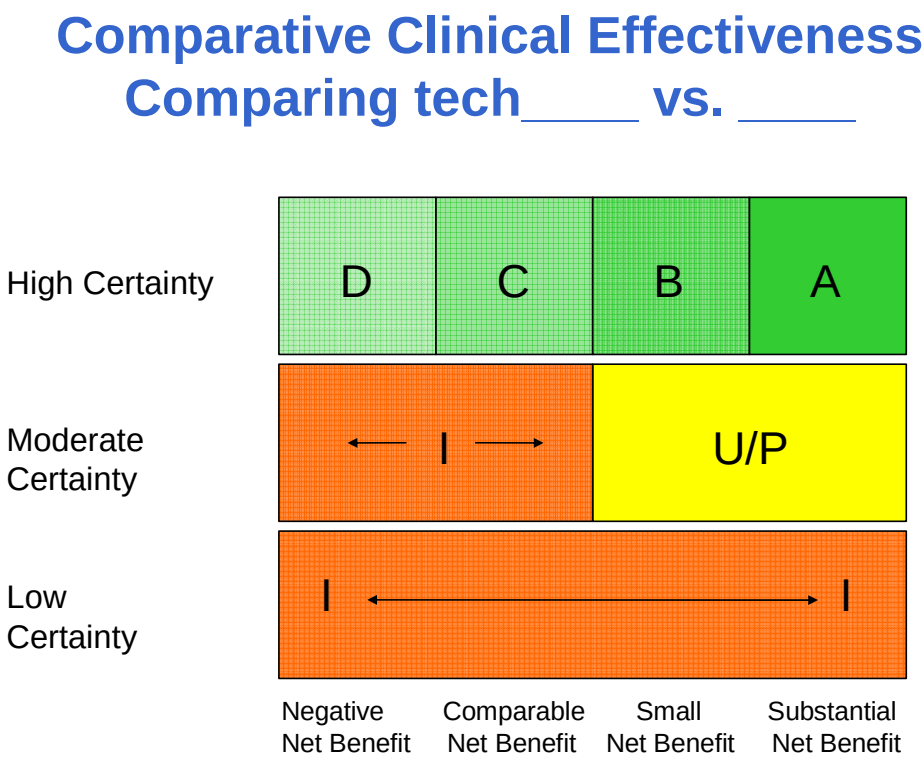
Richard Choo, MD, FRCPC, FACR Associate Professor of Radiation Oncology Mayo Clinic College of Medicine	None
Myriam Curet, MD, FACS Professor, Department of Surgery Stanford University Chief Medical Officer, Intuitive Surgical	Chief Medical Officer of Intuitive Surgical
Michele DiPalo Director, Health Services Evaluation Blue Cross & Blue Shield of Massachusetts	Employed by payer; involved in evaluation of new/emerging technology
Ted Ganiats, MD Chair, Department of Family & Preventive Medicine University of California at San Diego School of Medicine Executive Director, UCSD Health Services Research Center	No financial conflicts
G. Scott Gazelle, MD, MPH, PhD Director, Institute for Technology Assessment, Massachusetts General Hospital Professor of Radiology, Harvard Medical School Professor of Health Policy & Management, Harvard School of Public Health	None
Marthe Gold, MD Professor & Chair, Community Health and Social Medicine City College of New York	No financial conflicts, interested in strategies to trim healthcare budgets
Lou Hochheiser, MD Medical Director, Clinical Policy Development Humana, Inc.	Employed by payer; involved in clinical policy development
Jim C. Hu, MD, MPH Assistant Professor of Surgery, Harvard Medical School Director, Minimally Invasive Urologic Oncology Brigham & Women's Hospital	Perform open and robot-assisted prostatectomy
Phil Kantoff, MD Professor of Medicine Harvard Medical School Dana-Farber Cancer Institute	None

Terry Lindblom, MBA, PA-C Patient/Consumer Representative	Prostate cancer survivor
David Most, PhD Patient/Consumer Representative	Prostate cancer survivor
Lee Newcomer, MD Senior Vice President, Oncology UnitedHealthcare	Employed by payer
Catherine Tak Piech, MBA Vice President, Health Economics & Outcomes Research Centocor Ortho Biotech Services, LLC	Shareholder of Johnson & Johnson, which markets materials for minimally-invasive surgery through its Ethicon Endo-Surgery division
Alan B. Rosenberg, MD Vice President, Medical Policy, Technology Assessment, and Credentialing Programs Wellpoint, Inc.	Employed by payer, involved in coverage decisions; member, AHRQ Effective Health Care Stakeholder Group
Martin G. Sanda, MD Associate Professor of Surgery Harvard Medical School Beth Israel Deaconess Medical Center	Financial support from Beckman, Lilly, Amgen; involved in Phase II clinical trial of CyberKnife®
Ian Thompson, MD Professor and Chair, Department of Urology University of Texas HSC at San Antonio	Chair of AUA prostate cancer guideline committee; perform active surveillance and surgery; consultant to Veridex, a Johnson & Johnson company developing a biomarker for prostate cancer
Sean Tunis, MD, MSc Founding Director Center for Medical Technology Policy	No financial conflicts
David Veroff, PhD Vice President, Evaluation Services Health Dialog	Employed by patient information company

Methodology: ICER Integrated Evidence Rating™

Comparative Clinical Effectiveness

The ICER Integrated Evidence Rating™ combines a rating for comparative clinical effectiveness and a rating for comparative value. The clinical effectiveness rating arises from a joint judgment of the level of certainty provided by the body of evidence and the magnitude of the net health benefit -- the overall balance between benefits and harms. This method for rating the clinical effectiveness is modeled on the "Evidence- Based Medicine (EBM) matrix" developed by a multi-stakeholder group convened by America's Health Insurance Plans. This matrix is depicted below:



A = "Superior" [High certainty of a moderate-large net health benefit]
B = "Incremental" [High certainty of a small net health benefit]
C = "Comparable" [High certainty of a comparable net health benefit]
D = "Inferior" [High certainty of an inferior net health benefit]
U/P = "Unproven with Potential " [Moderate certainty of a small or moderate-large net health benefit]
This category is meant to reflect technologies whose evidence provides:

1) High certainty of *at least* comparable net health benefit
2) Moderate certainty suggesting a small or moderate-large net health benefit

I = "Insufficient" The evidence does not provide high certainty that the net health benefit of the technology is at least comparable to that provided by the comparator(s).

Certainty

The vertical axis of the matrix is labeled as a degree of certainty with which the magnitude of a technology's comparative net health benefit can be determined. This operational definition of certainty thus is linked to but is not synonymous with the overall validity, consistency, and directness of the body of evidence available for the assessment. ICER establishes its rating of level of certainty after deliberation by the Evidence Review Group, and throughout ICER follows closely the considerations of evidentiary strength suggested by the Effective Health Care program of the Agency for Health Research and Quality (AHRQ) (www.effectivehealthcare.org) and the GRADE working group (www.gradeworkinggroup.org).

High Certainty:

An assessment of the evidence provides high certainty in the relative magnitude of the net health benefit of the technology compared to its comparator(s).

Moderate Certainty:

There is moderate certainty in the assessment of the net health benefit of the technology. Moderate certainty implies that the evidence is limited in one or more ways so that it is difficult to estimate the net health benefit with precision. ICER's approach considers two qualitatively different types of moderate certainty. First, there may be limited certainty in the magnitude of any net health benefit, but there is high certainty that the technology is *at least* as effective as its comparator(s). The second kind of moderate certainty applies to those technologies whose evidence may suggest comparable or inferior net health benefit and for which there is not high certainty that the technology is at least comparable. These two different situations related to "moderate certainty" are reflected in the matrix by the different labels of "Unproven with Potential" and "Insufficient."

Limitations to evidence should be explicitly categorized and discussed. Often the quality and consistency varies between the evidence available on benefits and that on harms. We follow the GRADE and AHRQ approaches in highlighting key types of limitations to evidence, including:

a. Internal validity
    i. Study design
    ii. Study quality
b. Generalizability of patients (directness of patients)
c. Generalizability of intervention (directness of intervention)
d. Indirect comparisons across trials (directness of comparison)
e. Surrogate outcomes only (directness of outcomes)
f. Lack of longer-term outcomes (directness of outcomes)
g. Conflicting results within body of evidence (consistency)

Low Certainty:
There is low certainty in the assessment of net health benefit and the evidence is insufficient to determine whether the technology provides an inferior, comparable, or better net health benefit.

Net Health Benefit
The horizontal axis of the comparative clinical effectiveness matrix is "net health benefit." This term is defined as the balance between benefits and harms, and can either be judged on the basis of an empiric weighing of harms and benefits through a common metric (e.g., Quality Adjusted Life-Years, or "QALYs"), or through more qualitative, implicit weightings of harms and benefits identified in the ICER appraisal. Either approach should seek to make the weightings as explicit as possible in order to enhance the transparency of the ultimate judgment of the magnitude of net health benefit.

Whether judged quantitatively or qualitatively, there are two general situations that decision-making groups face in judging the balance of benefits and harms between two alternative interventions. The first situation arises when both interventions have the same types of benefits and harms. For example, two blood pressure medications may both act to control high blood pressure and may have the same profile of toxicities such as dizziness, impotence, or edema. In such cases a comparison of benefits and harms is relatively straightforward. However, a second situation in comparative effectiveness is much more common: two interventions present a set of trade-offs between overlapping but different benefits and harms. An example of this second situation is the comparison of net health benefit between medical treatment and angioplasty for chronic stable angina. Possible benefits on which these interventions may vary include improved mortality, improved functional capacity, and less chest pain; in addition, both acute and late potential harms differ between these interventions. It is possible that one intervention may be superior in certain benefits (e.g. survival) while also presenting greater risks for particular harms (e.g. drug toxicities). Thus the judgment of "net" health benefit of one intervention vs. another often requires the qualitative or quantitative comparison of different types of health outcomes.

Since net health benefit may be sensitive to individual patient clinical characteristics or preferences there is a natural tension between the clinical decision-making for an individual and an assessment of the evidence for comparative clinical effectiveness at a population level. ICER approaches this problem by seeking, through the guidance of its scoping committee, to identify a priori key patient subpopulations that may have distinctly different net health benefits with alternative interventions. In addition, the ICER appraisal will also seek to use decision analytic modeling to identify patient groups of particular clinical characteristics and/or utilities which would lead them to have a distinctly different rating of comparative clinical effectiveness.

The exact boundary between small and moderate-large net benefit is subjective and ICER does not have a quantitative threshold. The rating judgment between these two categories is guided by the deliberation of the Evidence Review Group.

Comparative Value

There are three categories of value: high, reasonable or comparable, and low. The ICER rating for comparative value arises from a judgment that is based on multiple considerations. ICER does not employ a single measure of cost-effectiveness for assignment of comparative value, nor does it rely on a formal threshold for determination of the level of value. Instead, comparative value is informed by multiple measures of potential economic impact, including:

- Impact on service use (e.g., tests, hospitalizations)
- Cost to reduce adverse outcomes (e.g., cost per hospitalization averted)
- Cost to achieve clinical success (e.g., cost per curative outcome)
- Cost per life year gained
- Cost per quality-adjusted life year (QALY) gained
- Budget impact per 1,000 diseased individuals
- System issues (e.g., manpower tradeoffs to invest in new technology)

The advantages for evaluating the full list of economic measures are twofold. First, the importance of these measures varies for individual stakeholders. For example, payers may be most interested in expressions of the clinical value achieved for the additional investment provided (e.g., cost per QALY, cost per event averted), while integrated health systems may ascribe most importance to measures of budgetary or system impact, and patients may be most interested in differential rates of downstream testing or other service use. Second, sole reliance on traditionally-accepted measures of cost-effectiveness such as cost per QALY may mask important considerations in evaluating whether to adopt a new technology. Cost-effectiveness findings may appear to be "reasonable" based on widely-used thresholds (e.g., $50,000 per QALY gained), when in reality the incremental investment required is for an imperceptible clinical gain.

ICER has developed a method for presenting multiple measures of economic impact together in a format known as the Comparative Value Evidence Table (CVET), which allows for visualization of economic measures important to each healthcare stakeholder. Wherever feasible, the CVET has been designed for interactive modification of certain economic model parameters and visualization of how findings might change. Uncertainty in model results is also explored through "sensitivity analyses"— analyses of the robustness of the economic model to changes in certain probabilities and/or costs. Assignment of comparative value is made based on the performance of the technology in question across all of these measures, in consultation with the ICER Evidence Review Group. An example of the summary table from the CVET can be found on the following page.

Details on the methodology underpinning the design and presentation of cost-effectiveness analyses within ICER appraisals are available on the ICER website at www.icer-review.org.

ICER Comparative Value Evidence Table (CVET)

Measure	Technology A	Technology B	Difference (B-A)
1. Service Impact			
Tests	27.4	17.9	(9.5)
Visits	31.6	24.8	(6.8)
Hospitalizations	0.0	1.0	1.0
Hospital days	0.0	3.0	3.0
Days of missed work	4.7	5.9	1.2
Pathway Total	63.7	52.6	(11.1)
2. Cost-Consequences			
$ to Prevent 1 Case of X		$210,000	
$ per Cure		$350,000	
3. Cost per Life-Year Saved		N/A	(equivalent survival)
4. Cost per QALY Gained		$1,050,000	
% of Cost/QALY <$100,000		2.63%	
SA 1: Surg Compl. 50% of Basecase		$547,000	
SA 2: ED 50% of Basecase		$442,000	
5. Budget Impact (per 1,000, 2 years)		$1,425,000	
6. Fixed Budget Tradeoffs		19.0	*Nurse FTEs @ $75K each*
		11.4	*MD FTEs @ $125K each*

Integrated Ratings

The ICER Integrated Evidence Rating™ combines the individual ratings given for comparative clinical effectiveness and comparative value. The overall purpose of the integrated ratings is to highlight the separate considerations that go into each element but to combine them for the purposes of conveying that clinical benefits provided by technologies come at varying relative values based on their cost and their impact on the outcomes of care and the health care system.

