RESEARCH ARTICLE
For reprint orders, please contact: reprints@futuremedicine.com

Using a population-based observational cohort study to address difficult comparative effectiveness research questions: the CEASAR study

Journal of **Comparative Effectiveness Research**

Daniel A Barocas*[1], Vivien Chen[2], Matthew Cooperberg[3], Michael Goodman[4], John J Graff[5], Sheldon Greenfield[6], Ann Hamilton[7], Karen Hoffman[8], Sherrie Kaplan[6], Tatsuki Koyama[1], Alicia Morgans[1], Lisa E Paddock[5], Sharon Phillips[1], Matthew J Resnick[1,9], Antoinette Stroup[10], Xiao-Cheng Wu[2] & David F Penson[1,9]

[1]Vanderbilt University, Nashville, TN, USA
[2]Louisiana State University, New Orleans, LA, USA
[3]University of California, San Francisco, San Francisco, CA, USA
[4]Emory University, Atlanta, GA, USA
[5]Cancer Institute of New Jersey – Robert Wood Johnson Medical School, New Brunswick, NJ, USA
[6]University of California, Irvine, Irvine, CA, USA
[7]University of Southern California, CA, USA
[8]MD Anderson Cancer Center, Houston, TX, USA
[9]Tennessee Valley Veterans Administration Health System, Nashville, TN, USA
[10]University of Utah, Salt Lake City, UT, USA
*Author for correspondence:
Tel.: +1 615 322 3807
Fax: +1 615 322 8990
dan.barocas@vanderbilt.edu

Background: While randomized controlled trials represent the highest level of evidence we can generate in comparative effectiveness research, there are clinical scenarios where this type of study design is not feasible. The Comparative Effectiveness Analyses of Surgery and Radiation in localized prostate cancer (CEASAR) study is an observational study designed to compare the effectiveness and harms of different treatments for localized prostate cancer, a clinical scenario in which randomized controlled trials have been difficult to execute and, when completed, have been difficult to generalize to the population at large. **Methods:** CEASAR employs a population-based, prospective cohort study design, using tumor registries as cohort inception tools. The primary outcome is quality of life after treatment, measured by validated instruments. Risk adjustment is facilitated by capture of traditional and nontraditional confounders before treatment and by propensity score analysis. **Results:** We have accrued a diverse, representative cohort of 3691 men in the USA with clinically localized prostate cancer. Half of the men invited to participate enrolled, and 86% of patients who enrolled have completed the 6-month survey. **Conclusion:** Challenging comparative effectiveness research questions can be addressed using well-designed observational studies. The CEASAR study provides an opportunity to determine what treatments work best, for which patients, and in whose hands.

KEYWORDS: active surveillance ■ comparative effectiveness research ■ observational study ■ prospective cohort study ■ prostate cancer ■ quality of life ■ radiation therapy ■ surgery

Management of localized prostate cancer is a high-priority field for primary comparative effectiveness research (CER) in the USA, for a number of key reasons [1–6]. Prostate cancer is the most commonly diagnosed noncutaneous malignancy in the USA, with approximately 240,000 new cases and 30,000 deaths per year [7]. While prostate cancer remains the second leading cause of cancer death among men in the USA, the vast majority of men diagnosed with prostate cancer have clinically localized disease, which can have a prolonged indolent phase, such that most will die of other causes, often before symptoms of prostate cancer manifest [8,9]. There are several acceptable management strategies for localized prostate cancer, broadly including surgery, radiation therapy (RT) and observation, among others [10,11]. Each of these treatments (acknowledging that observation is not really a treatment but a therapeutic strategy) has a unique effect on quality of life and cancer control, rendering management of localized prostate cancer highly preference sensitive [12–14]. To date, these important treatment decisions are made without the availability of adequate data upon which to compare the harms and benefits of the different treatments for specific patient groups, even as purveyors of newer modalities claim to yield superior results [15–18].

While the stakeholders of CER (patients, providers, payers and policy makers) often think first of randomized clinical trials (RCTs) as the optimal study design for

Future Medicine part of fsg

CER, other study designs, such as high-quality observational studies, may generate equally or perhaps even more valuable information upon which to base treatment decisions [5,19–22]. The example of localized prostate cancer illustrates many of the challenges to conducting RCTs and limitations in their applicability [4,23]. These include high cost; difficulty accruing patients willing to be randomized to different modalities of therapy; small sample size, narrow inclusion/exclusion criteria, and small number of racial and ethnic minority persons leading to poor generalizability; difficulty timing the trial (i.e., after the learning curve, but after the 'tipping point' of adoption); and difficulty accounting for the quality of the intervention, which is known to affect outcomes [24–27]. In short, RCTs are difficult to execute and, because of the focus on efficacy in a highly controlled setting, the results may not be broadly applicable.

The limitations of RCTs have been borne out in prostate cancer research designed to compare the effectiveness of surgery, radiation and observation. The European Medical Research Council (MRC) Prostate Cancer Working Party study known as PR06, set out to accrue 1800 patients, randomized to surgery, RT or observation, and closed after accruing only 35 patients in 2 years [28]. The Surgical Prostatectomy virus Interstitial Radiotherapy Intervention Trial (SPIRIT), a concerted effort among the American College of Surgeons Oncology Group (ACOSOG) and the National Cancer Institute of Canada, opened in 31 centers across North America in 2002, but closed after accruing only 56 patients in a 2-year period [29]. There have been two completed trials of surgery versus observation, the Prostate cancer Intervention Versus Observation Trial (PIVOT) and the Randomized Scandinavian Prostate Cancer Group Study Number 4 (SPCG-4). PIVOT screened 13,022 men with prostate cancer, approached 5023 and randomized 731 men (14.5% of eligible) from Veterans' Affairs Hospitals in the USA between 1994 and 2002 [30,31]. This fell far short of the 2000 intended patients, resulting in an inadequate sample size to address some of the important questions, such as the influence of risk stratum on comparative effectiveness of surgery versus observation. While the median age at randomization was only 67 years old, 48% of patients had died within a median follow-up of 10 years, suggesting that this is a much more infirm group of men than are typically considering radical surgery as a treatment for localized prostate cancer. The SPCG-4 accrued 695 men in Sweden, Finland and Iceland between 1989 and 1999 [32]. These were, for the most part, not screen-detected cancers as we most commonly see in practice currently. On the contrary, 75% of men had palpable disease and approximately half had a serum prostate specific antigen (PSA) greater than 10 ng/ml. Therefore, it is difficult to apply the results of the PIVOT and SPCG-4 trials to a contemporary population of men with varying degrees of baseline comorbidity and varying risk strata of prostate cancer. One promising RCT in prostate cancer treatment is the British Prostate Testing for Cancer and Treatment (ProtecT) trial, comparing surgery, radiation and observation, which accrued over 1600 patients (63% of eligible cases) between 2001 and 2009 [33,34]. Results are pending.

Observational study designs offer an alternative to RCTs [19–22]. The potential benefits of observational studies include the opportunity to evaluate treatment effectiveness (i.e., results as seen in practice) rather than efficacy (i.e., best-case scenario among a small homogenous group of RCT enrollees). The lower cost and lower barriers to accrual enable observational cohorts to enroll larger numbers of patients. As a result of this greater statistical power, it is possible to make inferences about differential effectiveness in subgroups of the population, and generalize results to the population at large. Of course, observational studies are subject to many sources of bias. Chief among these is confounding by indication, and the related concept of channeling, whereby patients with different baseline characteristics and disparate prognoses undergo different treatments, making it difficult to distinguish the results of treatment from the effect of baseline characteristics and pretreatment prognosis. Therefore, a well-designed observational study must define, *a priori*, clinically meaningful subgroups to permit data collection for known and proposed new confounders. Incorporating such data into statistical models that account for confounding secondary to nonrandom treatment assignment is a cornerstone of observational CER [35]. There is a track record of successful observational cohort studies using patient-reported outcomes in prostate cancer. However, these prior studies have suffered to some extent from the known limitations of this study design, which hamper our ability to use them to inform CER questions in this common malignancy. Some have been designed well, but

predate 21st century treatment modalities, such as robotic surgery, intensity-modulated RT and active surveillance [36]. Other studies have come from single institutions [37] or use a multicenter design that enrolls solely from tertiary referral academic medical centers, thereby limiting generalizability [38,39]. Others fail to account for the quality of the therapeutic intervention [40] or adequately capture baseline pretreatment functional or quality-of-life information [12]. Finally, none expand beyond the traditional clinical and sociodemographic variables to address some possible areas of residual confounding.

While both RCTs and observational studies are susceptible to obsolescing as secular trends alter the population of interest and/or treatments available, observational studies take far less time to accrue and, to that extent, may be more adaptable to such changes [25]. The effect of secular trends on relevance of study results may be particularly germane in prostate cancer CER, as therapies have evolved significantly over the past two decades (use of robotics in surgery, intensity-modulation and image guidance in radiation, and the use of active surveillance in favor of the passive watchful waiting approach to observation), even as the penetrance of screening has altered the stage of disease at presentation, and other factors may have altered the burden or reporting of baseline characteristics, such as comorbidity, urinary and sexual function.

In response to the considerable uncertainty surrounding the optimal treatment for localized prostate cancer, the Agency for Healthcare Research and Quality's (AHRQ) 2008 evidence report on the comparative effectiveness of therapies for localized prostate cancer concluded that "no one therapy can be considered the preferred treatment for localized prostate cancer due to the limitations in the body of evidence as well as the likely tradeoffs an individual patient must make between estimated treatment effectiveness, necessity and adverse effects" [41]. With this in mind, the AHRQ report calls for "high-quality, large prospective cohort studies … that identify men at the time of diagnosis and … collect comprehensive patient, tumor and treatment selection characteristics" [41]. In an effort to address this clarion call to action, we initiated a population-based observational cohort study entitled Comparative Effectiveness Analyses of Surgery and Radiation in localized prostate cancer, or CEASAR study. This study was funded by AHRQ through its Clinical and Health Outcomes Initiative in Comparative Effectiveness (CHOICE) award mechanism. This study stands as an example of how population-based observational studies that include rigorous methods and use tumor registries as a cohort inception tool can answer difficult CER questions in oncology. Given that many CER questions in cancer do not lend themselves easily to RCTs, CEASAR serves as a case study for future cancer-related CER.

The purposes of this paper are to describe the rationale and objectives of the CEASAR study and demonstrate how it is able to overcome some of the limitations of prior CER studies in this space; to document the design, methods and limitations of CEASAR; and to describe the baseline characteristics of this large population-based study of men with newly diagnosed, clinically localized prostate cancer. Doing so will provide a context for interpretation of the planned studies that will use the data generated by it.

Methods

Broadly, the aims of the CEASAR study are to determine which treatments for localized prostate cancer work, for which patients and in whose hands. Specifically, we will compare the effectiveness of different modalities of treatment, collecting baseline (pretreatment) health-related quality of life information followed by 6- and 12-month health-related quality of life data as the primary outcome. Secondly, we will identify patient-level characteristics (such as race, age, comorbidity, and baseline urinary and sexual function) that may influence the comparative effectiveness of specific management strategies. Lastly, we will use validated and proposed quality metrics to determine how comparative effectiveness of the various therapies varies by the quality of care received.

Our approach is to use a population-based, prospective cohort study design to assess the comparative effectiveness of contemporary management options for localized prostate cancer. Inclusion criteria are: men age under 80 years old with newly diagnosed, pathologically confirmed, clinically localized adenocarcinoma of the prostate diagnosed within 6 months of enrollment; PSA <50 ng/ml; English or Spanish speaking and; able to give consent. We accrued 3691 patients between January 2011 and February 2012; 3169 (85.9%) of whom have completed the 6-month survey. Although we are still

collecting 12-month outcome information, we are approximately 75% complete and have similar response rates compared to 6-month respondents at this time. A brief form of the protocol is available at [101] and the complete protocol has been submitted to AHRQ, as well as the local Internal Review Board (IRB).

■ CEASAR accrual methodology

We used a Rapid Case Ascertainment System (RCAS) to identify newly diagnosed cases of prostate cancer within five population-based Surveillance, Epidemiology and End Results (SEER) registry catchment areas (Atlanta, Los Angeles, Louisiana, New Jersey and Utah). More information on the SEER Registry is available at the website [102]. RCAS leverages electronic and written communication between pathology laboratories and the SEER registries in order to identify newly diagnosed cases rapidly. Thus, patients can be consented and baseline data can be gathered prior to initiation of therapy, or immediately thereafter. The SEER registries have been utilized successfully as a cohort inception tool in other studies [36] because this methodology minimizes selection bias, ensures that the cohort is a nationally representative sample and leverages the RCAS. In addition, the sites employ skilled investigators and professional abstractors who have demonstrated the ability to execute these studies, and yield high-quality data. Importantly, the inclusion of SEER sites in Atlanta, Louisiana, and Los Angeles ensured the accrual of a racially and ethnically diverse population, which is particularly important in the study of prostate cancer, since African–American men are disproportionately burdened by this disease [42]. The ability to over-sample from African–American communities serves to contrast this methodology with RCTs, which have difficulty accruing racial and ethnic minority subjects [24]. In addition, we accrued 262 patients from the Cancer of the Prostate Strategic Urologic Research Endeavor (CaPSURE™) observational disease registry, launched in 1995 and currently includes over 13,800 men diagnosed with prostate cancer at one of 40 community and academic practice sites nationwide [43]. By including CaPSURE patients, we intended to enrich the population with patients undergoing novel therapies and active surveillance, since many of the CaPSURE sites are large urology group practices, which tend to embrace emerging management strategies early on.

Of note, the central organizing site at Vanderbilt, each of the SEER sites and CaPSURE obtained IRB approval from the relevant local IRB. The diagnosing physician was contacted prior to contacting the patient, and the physician had the option of refusing consent to contact the patient. If consent to contact the patient was not refused, the patient was invited to participate. Returning the surveys constituted consent for participation. At the 6-month time point, each patient was required to sign consent for the 12-month medical record review.

We used a mixed-mode approach to the patient surveys to maximize response. At each time point, we mailed a survey to the patient's address of record in the registry. If the participant failed to respond the survey within 2–3 weeks, a member of the site research staff called the patient to determine if the study materials were received and answer any questions the subject may have had about the study. If necessary, a second survey was mailed. If the survey was not returned after another 2 weeks, a second reminder call was made. If, at either phone call, the patient preferred a telephone interview, it was completed at that time. The questionnaire was also translated into Spanish and interviews were conducted in Spanish when necessary. If a patient actively refused participation in the study, we did not attempt to collect further information and excluded him from the analysis. Extensive efforts were made to locate nonrespondents in order to achieve a high response rate. Additional methods used to trace individuals included internet search engines, Department of Motor Vehicles records and other public records.

■ Health-related quality of life survey

Questionnaires using existing, validated instruments were developed in conjunction with a psychometrician and content experts to collect patient-reported information at the baseline, 6-month and 12-month time points (see **Table 1** for a list of instruments and time points at which each was collected). These instruments assess domains including comorbidity [44], health-related quality of life [45–47], and disease-specific function and quality of life (e.g., sexual, urinary and bowel function) [48]. We also incorporated questionnaires to determine nontraditional patient-reported covariates and outcomes, including social support, level of engagement in healthcare decisions [49], impact of prostate cancer on emotion and function [50], prostate

Table 1. Patient-reported outcome measures.

Patient-reported measures	Instrument	Baseline patient survey	6-month patient survey	12-month patient survey
Race			×	
Age		×		
Income			×	
Education			×	
Marital status			×	
Employment status		×	×	×
Insurance			×	
Treatment received			×	
Family history of prostate cancer			×	
Cancer recurrence			×	×
Complications			×	×
Providers seen for prostate cancer care			×	×
Comorbidity	TIBI-CaP [44]		×	
General health	SF-36 [45–47]	×		
Emotional health, depression	SF-36, CES-D11 [45–47]	×	×	×
Prostate cancer-related anxiety	MAX-PC [51]		×	×
Functional limitations	SF-36, PF-10 [45–47,50]	×	×	×
Disease-specific HRQOL: urinary, sexual, bowel and hormonal	EPIC-26 [48]	×	×	×
Use of erectile aids		×	×	×
General HRQOL	SF-36 [45–47]	×	×	×
Impact of prostate cancer	PC Burden	×	×	×
Social support	SF-36, MOS Social Support Survey [45–47]	×		×
Illness management style	PDM scale [62], PDHCO scale [63]	×		×
Satisfaction with care			×	×

CES-D11: Center for Epidemiologic Studies Depression Scale; EPIC-26: Expanded Prostate Cancer Index Composite; HRQOL: Health-related quality of life; MAX-PC: Memorial Anxiety Scale for Prostate Cancer; MOS: Medical Outcomes Study; PC Burden: Prostate Cancer Burden scale; PDHCO: Provider-Dependent Health Care Orientation; PDM: Participatory decision-making; PF-10: Physical Functioning Subscale of the Short Form (36); SF-36: Short Form (36) Health Survey; TIBI-CaP: Total Illness Burden Index for Prostate Cancer.

cancer related anxiety [51], and satisfaction with care. Quality of life and functional status are the primary outcomes of the study, while the other domains are potential predictors or confounders. In addition, we asked for detailed information on demographics, including age, race, income, education, marital status and insurance. These data points, along with comorbidity, are important grouping variables for the second aim (i.e., to determine how patient level characteristics influence comparative effectiveness).

■ Medical record abstraction

At the 12-month time point, an extensive medical chart abstraction is being undertaken. The records are obtained from any urologist, radiation oncologist or medical oncologist who had a role in diagnosis, consultation or treatment of the prostate cancer, by asking the patient for a list of doctors and using available records to identify other providers involved in care. From these records, trained abstractors from the participating SEER sites enter data on diagnostic information (e.g., prebiopsy PSA, biopsy Gleason score, clinical stage, use of imaging for staging), treatment information (e.g., radiation target and dosing for RT patients, pathologic findings in surgery patients), follow-up (e.g., post-treatment PSAs, additional treatments) and treatment-related complications. We also

abstract a number of quality-of-care indicators derived from Physician Quality Reporting Initiative (PQRI) and Research and Development (RAND) Corporation (Table 2) [52,53].

The methodology for obtaining information from the medical record is similar to the protocol for data abstraction used for creation of the SEER registry itself, and has been used successfully in numerous prior studies using SEER sites as cohort inception tools. Abstractor training was conducted in a series of face-to-face and web-based conferences, followed by monthly phone calls throughout the data collection period. A series of exercises were conducted prior to completing abstractor training to ensure standardization of abstraction procedures. Each site was required to double-abstract 3–5% of all cases to evaluate for inter-rater reliability of key abstracted items. This information was fed back to the sites to correct any systematic abstraction issues, and it was used to eliminate items with poor inter-rater reliability.

Data management

Data were collected by each site and gathered at the central organizing site in Nashville (TN, USA). Questionnaires were then verified, cleaned, scanned and merged into electronic analytic files. The medical charts were abstracted directly into Vanderbilt's Research Electronic Data Capture (REDCap) system, a secure web-based data repository, specifically designed to support data capture for research studies [54]. REDCap is ideally suited for a multisite study of this type because of its intuitive interface, forms that are easy to create and entirely customizable, ability for remote entry, differential access rights for each user or user group (so that one can view only certain cases), and output formats that are compatible with all commonly used statistical packages. All data delivered to the central site are de-identified.

Table 2. Selected quality-of-care indicators included in CEASAR.

Selected quality-of-care indicators	Source	Ref.
Avoidance of bone scan in low-risk patients	PQRI	[53]
Documentation of pretreatment rectal examination, clinical stage, PSA and biopsy Gleason score	RAND	[52]
Documentation of baseline function by provider	RAND	[52]
Communication with primary care provider by treating provider	RAND	[52]
Documentation of counseling regarding all treatment options	RAND	[52]
CT planning (for XRT patients)	RAND	[52]
Radiation dose received (for XRT patients)	RAND	[52]
Receipt of 3D conformal beam radiation (for XRT patients)	PQRI	[53]
Use of ADT for high-risk patients (for XRT patients)	PQRI	[53]
Report of Gleason grade following surgery	RAND	[52]
Report of margin status following surgery	RAND	[52]
Report of pathologic stage following surgery	RAND	[52]

ADT: Androgen deprivation therapy; CT: Computed tomography; PQRI: Physician Quality Reporting Initiative; PSA: Prostate-specific antigen; RAND: Research and development; XRT: External beam radiation therapy.

Statistical methods

The study is powered to identify differences on multivariable regression in outcome (12-month quality of life, measured by scores on the Expanded Prostate Cancer Index Composite [EPIC] domains) between patients undergoing surgery and those undergoing radiation. Secondary outcomes of interest include complications of treatment, and cancer control (although the number of events at 12 months after diagnosis is predicted to be too low to allow for meaningful comparisons, and we hope to procure additional funding to study these end points at longer follow-up). We expect to have power to identify differences in quality of life between different types of surgery (open and robotic) and types of radiation (brachytherapy, traditional external beam RT and intensity-modulated RT); and to determine whether outcomes differ within treatment group by patient-level variables (race/ethnicity, education level and income). Propensity score matching between surgery and radiation groups will be used to address residual confounding given the nonrandom treatment assignment. Included in the propensity score model will be age at diagnosis, study center, baseline PSA, Gleason Score, clinical stage, comorbidity score, income, education level, race, and baseline sexual, urinary and bowel function. We will also use the quality metrics as predictors of quality of life to determine the impact of physician performance on the outcomes of treatment.

A unified, generalized regression approach will be used to evaluate the association between explanatory variables and outcomes. Based on the types of response variables, we will use logistic regression (dichotomous), ordinal logistic regression (ordered categories), ordinary regression (numerical), Cox proportional hazard regression (time to event), and mixed-effects model (repeated measurements). For the first aim, the explanatory variable of interest will be treatment choice (surgery vs radiation), while the second aim will focus on patient characteristics (race/ethnicity, comorbidity score, education level and

household income). The third aim, the quality-of-care analysis, will evaluate provider performance as a predictor of outcome, which may entail the creation of a composite quality measure, because single-item measures of quality are less reliable that multi-item measures for assessing performance at different levels of the healthcare system (e.g., physician, site).

For all analyses, continuous variables will be allowed to have nonlinear associations with the response variable by the use of regression spline. Model validation will be conducted through bootstrap method that will allow estimation of the likely future performance of a predictive model, and adjustment of the regression coefficients to account for over-fitting. Based on predictions of the number of patients in each treatment group, we estimated that a minimum total sample size of 2500 patients would be required to detect a difference between groups of half of a standard deviation in scores on the sexual and urinary function domains of the EPIC instrument, with approximately 90% power and a type I error rate of 5%.

Results

Response rate

We accrued 3691 patients from five SEER sites and the CaPSURE registry in a 14-month period. The response rate among eligible persons was 50.2% (Figure 1). A total of 85.9% of eligible participants who completed a baseline survey have returned the 6-month survey; 12-month data collection is ongoing.

Representativeness

The mean age of the population at accrual was 64.8 (median: 65; range: 40–80 years), which mirrors national statistics (median age at diagnosis in SEER: 67) and highlights the need to study cohorts that are not based on the Medicare population [103]. Twenty-four percent of CEASAR participants self-identified as nonwhite, which is somewhat higher than the 17.5% nonwhite incident prostate cancer patients in SEER between 2001 and 2007 [55]. Approximately 3% of participants have filled out the Spanish language version of the baseline or 6-month survey. Forty percent of patients had a household income under US$50,000 (median household income in the USA in 2007–2011: $52,762) [104]. Nineteen percent had no insurance or were inadequately insured, which is consistent with national statistics for 2010 (16.3% of the USA population uninsured) [105].

Fifty-one percent of 6-month respondents who had already selected a treatment reported surgery as their primary therapy, while 30% reported radiation and 11% reported active surveillance. We will use the medical chart abstraction data to obtain further detail on primary therapy, such as distinguishing between subtypes of RT and surgery, and defining watchful waiting as the group who received no therapy, but did not explicitly select active surveillance. Table 3 shows baseline demographics for the 2009 patients who had completed a 6-month survey and reported selection of one of these three management options. Not surprisingly, these findings demonstrate significant differences among groups in baseline characteristics. As indicated above, statistical modeling, including propensity score and multivariable analysis will be essential in evaluating the association between treatment and outcome.

The availability of self-reported treatment choice enables us to perform preliminary analyses of predictors of treatment choice, to inform later models of the association between treatment and downstream outcomes. Preliminary univariate and multivariable analyses indicate that health status, as measured by Total Illness Burden Index for Prostate Cancer (TIBI-Cap) and the overall health scale from the Short Form (36) Health Survey (SF-36), as well as advanced age influence treatment choice (Table 4). As these factors are also expected to influence response to treatment, their inclusion in final models of the association between treatment and quality of life outcome will be essential.

In addition to demographic and comorbidity data, we measured baseline disease-specific function using the EPIC-26, which has several domains, each scaled from 0 to 100, with higher scores indicating better function (Table 4). Responses to individual items are also pertinent to treatment decision and outcomes. For example, 28% of participants reported some degree of pretreatment urinary incontinence, defined as urinary leakage at least once a week. However, only 7–8% of reported using pads for incontinence at baseline and 8% of participants reported moderate or severe bother secondary to urinary incontinence. Nineteen percent of men reported moderate-to-severe bother secondary to the subjective sensation of weak stream or incomplete emptying and 24% reported moderate-to-severe bother secondary to urinary frequency. Pretreatment erectile dysfunction was common among men with newly diagnosed

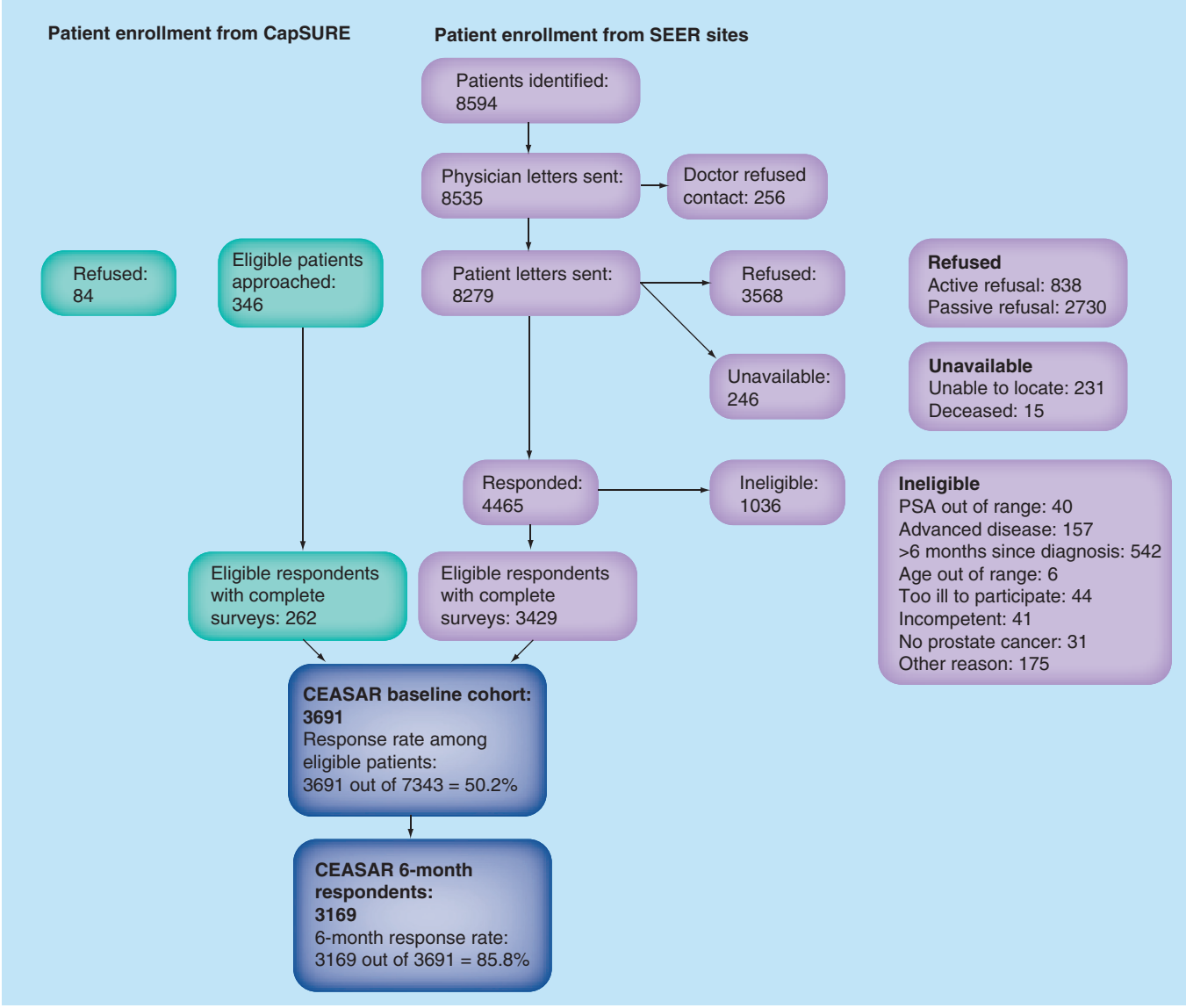

Figure 1. CEASAR cohort composition.
CapSURE: Cancer of the Prostate Strategic Urologic Research Endeavor; PSA: Prostate-specific antigen; SEER: Surveillance, Epidemiology and End Results.

prostate cancer, with 48% of participants reporting an inability to achieve erection sufficient for intercourse. There was little pretreatment bowel dysfunction, reflected in the mean bowel function domain summary score of 93. Urinary, sexual and bowel domain function scores differed significantly between those choosing surgery and those selecting radiation. Again, these findings not only provide a baseline assessment of pretreatment function, they are also important confounders of the relationship between treatment and quality of life outcome, which have been lacking in prior population-based studies.

In addition to sociodemographic information, comorbidity and baseline quality of life, we have accumulated a wealth of baseline self-reported data on nontraditional scales, such as engagement in treatment decisions, social support and anxiety about prostate cancer. Preliminary analyses confirm that these may explain some of the residual confounding in prior studies. For example, we find lower participation in medical decision-making among nonwhite participants and show that higher levels of participation are independently associated with treatment choice. Not only do these findings indicate a potential

target for quality improvement among nonwhite patients, they also identify participatory decision-making as a confounder of the relationship between race and treatment choice. The latter will factor into our modeling for the determinants of outcome in racial and ethnic minority subgroups.

The medical chart review at the 12-month time point will be used to confirm self-reported treatment choice and to collect copious details

Table 3. Patient demographics.

Characteristic	Active surveillance (n = 236); % (n)	Surgery (n = 1116); % (n)	Radiation therapy (n = 657); % (n)
Age, years (p < 0.001)			
<60	22 (51)	40 (447)	16 (102)
60–64.9	21 (50)	26 (291)	16 (105)
65–69.9	26 (62)	22 (241)	28 (181)
70+	31 (73)	12 (135)	41 (267)
Race (p < 0.001)			
White/Caucasian	78 (185)	75 (831)	72 (466)
African–American	11 (27)	12 (137)	18 (119)
Hispanic/Latino	6 (15)	8 (94)	6 (41)
Other	4 (9)	4 (49)	4 (23)
Insurance (p < 0.001)			
Private/HMO	40 (95)	56 (621)	27 (175)
Medicare only	41 (96)	24 (261)	42 (273)
Veterans'/military	6 (14)	4 (44)	8 (55)
Underinsured or uninsured	13 (30)	16 (181)	23 (149)
Employment status (p < 0.001)			
Full-time	35 (83)	50 (558)	24 (157)
Part-time	7 (16)	7 (82)	6 (40)
Retired	57 (134)	37 (407)	65 (424)
Unemployed	1 (3)	5 (60)	4 (29)
Income (p < 0.001)			
<US$30,000	13 (30)	18 (192)	29 (173)
US$30,001–50,000	22 (49)	17 (182)	22 (131)
US$50,001–100,000	32 (72)	33 (349)	30 (181)
>US$100,000	33 (76)	31 (321)	20 (118)
Education (p < 0.001)			
Some high school, technical school or less	7 (16)	9 (102)	13 (84)
High school or technical school graduate	13 (31)	21 (234)	21 (139)
Some college	21 (50)	22 (240)	26 (170)
College graduate	23 (55)	25 (273)	21 (139)
Graduate or professional school	36 (84)	24 (262)	19 (121)
Marital status (p = 0.006)			
Never married	3 (7)	5 (57)	6 (37)
Married	80 (188)	82 (908)	75 (491)
Separated, divorced, widowed	17 (41)	13 (148)	19 (126)

HMO: Health maintenance organization.

Table 4. Premorbid health status, quality of life and disease-specific function by treatment choice.

Measure	Active surveillance (n = 236)	Surgery (n = 1116)	Radiation therapy (n = 657)	p-value
General health				
TIBI-Cap	3 (2–5)	3 (2–4)	4 (3–5)	<0.001
Quality of life				
SF-36				
General health:				
■ 0	0% (0)	0% (3)	0% (3)	
■ 20	2% (4)	1% (11)	1% (7)	
■ 40	6% (14)	5% (58)	13% (87)	
■ 60	29% (68)	27% (298)	37% (242)	
■ 80	43% (102)	43% (481)	35% (230)	
■ 100	20% (48)	24% (262)	13% (87)	
Physical function	95 (85–100)	100 (85–100)	90 (75–100)	<0.001
Mental function	84 (72–92)	84 (68–92)	88 (72–92)	0.001
Energy/vitality	75 (65–85)	75 (60–85)	75 (60–85)	0.016
Disease-specific function (p < 0.001)				
EPIC				
Incontinence	100 (86–100)	100 (79–100)	100 (81–100)	0.562
Urinary irritation	88 (75–94)	88 (75–100)	88 (75–94)	0.874
Sexual function	75 (44–92)	75 (40–96)	58 (26–83)	<0.001
Bowel function	100 (96–100)	100 (92–100)	100 (92–100)	0.064

SF-36 and EPIC scores are scaled from 0 to 100, with higher scores indicating better quality of life or better function. The SF-36 general health is a single-item measure, so it is presented categorically.

Values are represented as medians (25–75th percentile) or percentages (n).

EPIC: Expanded Prostate Cancer Index Composit; SF-36: Short Form (36) Health Survey; TIBI-Cap: Total Illness Burden Index for Prostate Cancer.

of diagnosis, treatment and quality of care. This will enable us to control for the quality of the intervention in assessing the association between treatment and outcome. For example, rather than merely classifying the patient as having undergone radiation, we will be able to specify the type of RT used, the dose received, the target of RT (prostate, seminal vesicles and/or lymph nodes), use of image guidance, among others. Furthermore, linkage to provider datasets will enable us to control for provider characteristics as well.

Comment

The CEASAR study is a robust response to the AHRQ's call for additional data to inform decision-making in localized prostate cancer. The use of a well-designed prospective cohort study to address CER priorities in prostate cancer could serve as a model for other disease states as well. The constraints of the funding mechanism have limited the initial study to 12 months of follow-up time, which, in light of the high long-term survival rates for localized prostate cancer, is clearly insufficient to allow for a complete comparison of the effectiveness of different treatments. Therefore, the intention is to obtain further funding to extend the length of follow-up for patient-reported outcomes (PROs) and cancer control outcomes. The features of the CEASAR study that contribute to its value and, more broadly, support the use of observational studies to address complex CER questions, are enumerated below.

■ Study design

Well-designed prospective cohort studies may be the optimal study design for CER in localized prostate cancer. Preclinical studies and clinical trials are necessary steps in identifying new treatments an evaluating their efficacy in isolation. However, while RCTs have advantages in terms of minimizing bias, isolating the treatment

effect and optimizing internal validity, they are problematic for several reasons. First, the cost of high-quality RCTs is frequently prohibitive and randomizing patients in the USA to different treatment modalities has proven difficult. Second, because of small sample size and poor accrual of racial minority persons, the differential impact of the intervention on subgroups of patients cannot be determined, thereby limiting generalizability. Similarly, because of strict inclusion/exclusion criteria, they demonstrate the efficacy of the treatment in a very narrow context, thus limiting the external validity. Trials also experience difficulty controlling for quality of the intervention. Finally, the length of time required to recruit sufficient numbers of patients for RCTs presents special challenges in fields influenced by rapid uptake of new technologies. For example, the PIVOT and SPCG-4 trials each took 8–10 years to accrue 700 patients, an interval long enough to include the 'tipping point' of a new technology, after which patients may become very resistant to randomization [25]. By contrast, a prospective cohort study allows for determination of treatment effectiveness (i.e., results seen in real-world practice) and has sufficient sample size to generate meaningful estimates of the effectiveness of treatment in specific subgroups of the population. Both of these factors make the results of observational studies more generalizable and allow for the opportunity to truly compare the effectiveness of different treatments as they are delivered and in the population likely to use those treatments. Furthermore, the rapid accrual of patients makes observational studies potentially less susceptible to secular trends. Observational studies are always subject to confounding by indication and other bias, but these can be minimized to an extent by collecting data on known and hypothesized confounders and by using adjustment techniques such as propensity score and instrumental variable analysis to account for residual confounding [36,56].

Accrual mechanism

One of the key concerns in interpreting observational studies is that the selection of a study population be free of bias and be representative of the population at risk. Using the RCAS via SEER registry sites, along with the CaPSURE accrual mechanism as cohort inception tools has enabled the CEASAR study to rapidly accrue a large, nationally representative, diverse sample of men with newly diagnosed prostate cancer. The RCAS identifies patients shortly after diagnosis, allowing for collection of baseline (pretreatment) self-reported quality of life and function, without which subsequent patient-reported outcomes would be difficult to interpret. In addition, because of the size and diversity of the cohort, the study will be applicable to individual patients by evaluating the effect of important patient-level characteristics, such as race, socioeconomic status and comorbidity conditions on comparative effectiveness. In addition, the rapid accrual of a large study cohort permits evaluation of emerging technologies in a timely and relevant fashion.

Focus on PROs as the primary outcomes

The 5-year survival for localized prostate cancer approaches 100% [57], and survival remains high even in the long term, without evidence of a substantial benefit of one treatment compared to others. Therefore, prostate cancer treatment is uniquely preference-sensitive. Thus, we elected to focus on outcomes that matter most to patients: quality of life, post-treatment function, and complications of treatment. In this way, our study is truly patient-centered. The results will help patients and their doctors to align management of localized prostate cancer with patient values and preferences with regard to these important outcomes.

Risk adjustment

In order to minimize bias and confounding, nonrandomized studies must be particularly attentive to risk adjustment. The extensive patient surveys and medical chart review in the CEASAR study enabled us to control for a multitude of factors that may influence outcomes. In addition to collecting known covariates in many domains, we have included nontraditional patient-centered scales that may influence both choice of treatment and treatment outcome, such as engagement in participatory decision-making. As such, these psychosocial variables meet the 'classic' definition of a confounding variable. Observational studies in prostate cancer do not routinely capture these types of variables and, to this end, we believe that the inclusion of these nontraditional scales will result in much improved risk adjustment in CEASAR. Furthermore, the *a priori* specification of treatment groups and the use of propensity score matching are designed to minimize bias in the setting of a nonrandomized study [5].

■ Focus on harms & benefits of treatment alternatives in real-world practice

It is critical for a CER study to account for the heterogeneity of the patient, the disease, the provider and the intervention to determine not only what treatment works best on average but what works best for which patient, and in whose hands [5,58,59]. In the CEASAR study, patients can be categorized according to comorbidity burden, baseline function, race, socioeconomic status, and so forth, in order to determine what treatment works best for which patients. We will obtain disease-specific information on PSA, clinical stage and clinical Gleason score in order to risk-stratify patients. The relatively large sample size gives us ample power to compare the effectiveness of subtypes of treatments (e.g., open and robotic surgery) and to evaluate differential response to therapy among different patient groups (e.g., racial groups, socioeconomic status).

■ Focus on quality of care

RCTs and observational studies alike have been critiqued for failing to account for the quality of the intervention, particularly in situations like a surgical intervention, where quality of care appears to impact outcomes. Although there are ample data to demonstrate a relationship between quality of care and outcomes in the treatment of prostate cancer, no previous comparative effectiveness study in prostate cancer has attempted to control for the quality of the intervention [26,52,60,61]. The CEASAR study is unique in this regard and we anticipate that patients will be able to use our results to guide not only treatment decisions, but also provider and facility decisions.

Broadly, we are generating and disseminating data that patients and providers can use to determine what treatment works best, for which patients, and in whose hands. These central comparative effectiveness findings will be available after 12-month data collection is complete in mid-2013, and will fill knowledge gaps recognized by AHRQ, Patient-Centered Outcomes Research Institute, Institute of Medicine and the NIH, and, more importantly, by prostate cancer patients, their loved ones, and their doctors.

Additionally, the dataset offers opportunities beyond the main study aims. For example, we have begun to explore associations between baseline characteristics (such as sociodemographics, comorbidities, quality of life, function, engagement in participatory decision-making) and treatment choice. Furthermore, we have begun to compare the baseline characteristics of patients in the CEASAR cohort to those in a similarly accrued cohort from the mid-1990s, the Prostate Cancer Outcomes Study (PCOS). By doing so, we may be able to make inferences about changes in the recognition of functional deficits (e.g., erectile dysfunction, urinary incontinence) across eras. Other opportunities for this dataset include linkage with provider and facility information in order to determine how structural quality indicators (e.g., provider case volume, hospital bed size) influence outcomes. Finally, we may be able to link CEASAR with payer information compare cost of different treatments as well.

Conclusion

There are several options for study design for CER, including RCTs and well-designed prospective cohort studies, each of which has its strengths and limitations. For CER in localized prostate cancer, randomized trials have proven difficult to execute and the results of those that have been completed have been difficult to generalize to the broader population. For these reasons, prospective cohort studies may be better suited to answer those questions most germane to CER in localized prostate cancer.

By utilizing a prospective cohort study design, being population based, collecting data on quality of care and nontraditional scales, collecting baseline patient-reported functional status and quality of life, the CEASAR study is poised to answer the critical questions in localized prostate cancer (i.e., what treatment works best, for which patients and in whose hands). This information will be valuable to patients and providers who are faced with treatment decisions, and will provide a resource for investigators, complementing past and pending RCT results.

AHRQ's investment in the CEASAR study will pay dividends in the near term, as 12-month outcomes are collected. However, post-treatment function and its impact on quality of life continues to evolve for many years after diagnosis, and cancer control outcomes may not be apparent for a decade or more. Thus, we suspect that the benefits of AHRQ's investment in the CEASAR cohort will continue to accrue for 10–15 years or more, as long as funding for CER remains available.

Future perspective

While RCTs have long been viewed as the standard for CER, the challenges to their execution

(e.g., expense, difficulty accruing patients) and inherent limitations (e.g., lack of generalizability, sample size insufficient to perform subgroup analyses) highlight the need for other options. Well-designed prospective observational cohort studies offer one such alternative. The observational design allows for more rapid and efficient accrual, yielding sufficient sample size and heterogeneity to evaluate how variation in baseline characteristics may influence the outcome. CER in prostate cancer has benefited from observational studies over the past two decades, but the CEASAR study offers the promise of in-depth comparison of contemporary treatment options for prostate cancer, while accounting for the substantial heterogeneity of the population and the quality of intervention. The CEASAR study will yield important information to guide treatment decisions in localized prostate cancer, and will also serve as a model of the value of observational studies in CER. Further funding may provide opportunities to study long-term PROs, as well as cancer control in order to develop a more complete picture of the comparative effectiveness of different treatments for localized prostate cancer.

Financial & competing interests disclosure

This study was supported by the Agency for Healthcare Research and Quality (1R01HS019356). The data management was facilitated by the use of Vanderbilt University's Research Electronic Data Capture (REDCap) system, which is supported by the Vanderbilt Institute for Clinical and Translational Research grant (UL1TR000011

Executive summary

Background

- Randomized controlled trials in prostate cancer have proven difficult to execute and, when executed, difficult to generalize to the population at risk.
- Prospective cohort studies offer an alternative to randomized controlled trials in performing comparative effectiveness research.
- Potential benefits include lower cost, more efficient accrual, greater number of patients, ability to control for heterogeneity of patient characteristics and quality of care.
- The yield is an estimate of how different interventions compare under real-world circumstances (i.e., true comparative effectiveness).
- The Comparative Effectiveness Analyses of Surgery and Radiation (CEASAR) study in localized prostate cancer responds to the lack of data regarding comparative effectiveness of different treatment options for localized prostate cancer. In doing so, it answers the Agency for Healthcare Research and Quality's call for well-designed prospective cohort studies in this population and addresses an Institute of Medicine priority for comparative effectiveness research.

Methods

- The aims of the study are to determine what treatments result in the most favorable outcomes, as measured by complications, quality of life and disease-specific function. In addition, the study aims to determine what works best in prespecified subgroups of men, and how the quality of care influences the association between treatment and outcome.
- The accrual methodology utilizes cancer registries as cohort inception tools and harnesses the resources of the registry sites to identify cases, enroll patients, collect patient-reported outcomes and clinical information.
- An extensive health-related quality of life instrument, comprised of validated instruments is administered at baseline, 6 months, and 12 months after enrollment.
- The patient's medical chart is abstracted at 12 months in order to obtain disease characteristics, details of treatments and complications, and to assess quality-of-care measures.
- Data are managed centrally, using a secure web-based data repository, known as Research Electronic Data Capture (REDCap).
- We plan to use a unified, generalized regression approach to evaluate the association between explanatory variables (treatment) and outcomes (scores on quality-of-life measures). A minimum of 2500 patients is required to demonstrate a clinically and statistically significant difference in outcome between surgery and radiation.

Results

- The CEASAR study has accrued 3691 men with newly diagnosed prostate cancer, collecting baseline demographic and quality-of-life information, in addition to some nontraditional scales that may prove to account for some of the residual confounding seen in prior studies. An extensive chart review is under way to collect clinical information and quality-of-care metrics.
- The CEASAR cohort is representative of the national demographics of prostate cancer patients.
- Preliminary results demonstrate potentially important differences in baseline characteristics and quality of life between patients who chose active surveillance, surgery and radiation.

Conclusion

- The CEASAR study provides a rich resource for comparative effectiveness research in prostate cancer and serves as a model for observational studies in comparative effectiveness research.

from NCATS/NIH). The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Ethical conduct of research

The authors state that they have obtained appropriate institutional review board approval or have followed the principles outlined in the Declaration of Helsinki for all human or animal experimental investigations. In addition, for investigations involving human subjects, informed consent has been obtained from the participants involved.

References

Papers of special note have been highlighted as:

- ■ of interest
- ■■ of considerable interest

1 Institute of Medicine. *Initial National Priorities for Comparative Effectiveness Research.* The National Academies Press, DC, USA (2009).

2 Sox HC. Defining comparative effectiveness research: the importance of getting it right. *Med. Care* 48(Suppl. 6), S7–S8 (2010).

3 Iglehart JK. Prioritizing comparative-effectiveness research – IOM recommendations. *N. Engl. J. Med.* 361(4), 325–328 (2009).

■ **Outlines the Institute of Medicine's priorities for comparative effectiveness research.**

4 Wang AT, Wang JK, Montori VM, Murad MH. Comparative effectiveness research in urology. *World J. Urol.* 29(3), 277–282 (2011).

5 Greenfield S, Kaplan SH. Building useful evidence: changing the clinical research paradigm to account for comparative effectiveness research. *J. Comp. Eff. Res.* 1(3), 263–270 (2012).

■ **Cogent review of the evolving paradigm for comparative effectiveness research.**

6 Sox HC, Greenfield S. Comparative effectiveness research: a report from the Institute of Medicine. *Ann. Intern. Med.* 151(3), 203–205 (2009).

7 Siegel R, Naishadham D, Jemal A. Cancer statistics, 2013. *CA Cancer J. Clin.* 63(1), 11–30 (2013).

8 Shao YH, Demissie K, Shih W *et al.* Contemporary risk profile of prostate cancer in the United States. *J. Natl Cancer Inst.* 101(18), 1280–1283 (2009).

9 Albertsen PC, Hanley JA, Fine J. 20-year outcomes following conservative management of clinically localized prostate cancer. *JAMA* 293(17), 2095–2101 (2005).

10 Thompson I, Thrasher JB, Aus G *et al.* Guideline for the management of clinically localized prostate cancer: 2007 update. *J. Urol.* 177(6), 2106–2131 (2007).

11 Mohler J, Bahnson RR, Boston B *et al.* NCCN clinical practice guidelines in oncology: prostate cancer. *J. Natl Compr. Canc. Netw.* 8(2), 162–200 (2010).

12 Penson D, McLerran D, Feng Z *et al.* 5-year urinary and sexual outcomes after radical prostatectomy: results from the prostate cancer outcomes study. *J. Urol.* 173(5), 1701–1705 (2005).

■■ **Demonstrates the power of using well-designed prospective cohort studies to compare the effectiveness of different treatments for localized prostate cancer.**

13 Steineck G, Helgesen F, Adolfsson J *et al.* Quality of life after radical prostatectomy or watchful waiting. *N. Engl. J. Med.* 347(11), 790–796 (2002).

14 Ramsey SD, Zeliadt SB, Fedorenko CR *et al.* Patient preferences and urologist recommendations among local-stage prostate cancer patients who present for initial consultation and second opinions. *World J. Urol.* 29(1), 3–9 (2011).

15 Kang DC, Hardee MJ, Fesperman SF, Stoffs TL, Dahm P. Low quality of evidence for robot-assisted laparoscopic prostatectomy: results of a systematic review of the published literature. *Eur. Urol.* 57(6), 930–937 (2010).

16 Joseph JV, Vicente I, Madeb R, Erturk E, Patel HRH. Robot-assisted vs pure laparoscopic radical prostatectomy: are there any differences? *BJU Int.* 96(1), 39–42 (2005).

17 Lips I, Dehnad H, Kruger AB *et al.* Health-related quality of life in patients with locally advanced prostate cancer after 76 Gy intensity-modulated radiotherapy vs. 70 Gy conformal radiotherapy in a prospective and longitudinal study. *Int. J. Radiat. Oncol. Biol. Phys.* 69(3), 656–661 (2007).

18 Ramsay C, Pickard R, Robertson C *et al.* Systematic review and economic modelling of the relative clinical benefit and cost-effectiveness of laparoscopic surgery and robotic surgery for removal of the prostate in men with localised prostate cancer. *Health Technol. Assess.* 16(41), 1–313 (2012).

19 Yang W, Zilov A, Soewondo P, Bech OM, Sekkal F, Home PD. Observational studies: going beyond the boundaries of randomized controlled trials. *Diabetes Res. Clin. Pract.* 88(Suppl. 1), S3–S9 (2010).

20 Lohr KN. Comparative effectiveness research methods: symposium overview and summary. *Med. Care* 48(Suppl. 6), S3–S6 (2010).

21 Norris SL, Atkins D, Bruening W *et al.* Observational studies in systemic reviews of comparative effectiveness: AHRQ and the effective health care program. *J. Clin. Epidemiol.* 64(11), 1178–1186 (2011).

22 Concato J, Lawler EV, Lew RA, Gaziano JM, Aslan M, Huang GD. Observational methods in comparative effectiveness research. *Am. J. Med.* 123(12), E16–E23 (2010).

23 Lavallée LT, Fergusson D, Breau RH. The role of randomized controlled trials in evidence-based urology. *World J. Urol.* 29(3), 257–263 (2011).

24 Al-Refaie WB, Vickers SM, Zhong W, Parsons H, Rothenberger D, Habermann EB. Cancer trials versus the real world in the United States. *Ann. Surg.* 254(3), 438–443 (2011).

25 Schroeder TV. Evidence-based medicine in rapidly changing technologies. *Scand. J. Surg.* 97(2), 100–104 (2008).

26 Barocas DA, Mitchell R, Chang SS, Cookson MS. Impact of surgeon and hospital volume on outcomes of radical prostatectomy. *Urol. Oncol.* 28(3), 243–250 (2010).

27 Begg CB, Riedel ER, Bach PB *et al.* Variations in morbidity after radical prostatectomy. *N. Engl. J. Med.* 346(15), 1138–1144 (2002).

28 PR06 Collaborators. Early closure of a randomized controlled trial of three treatment approaches to early localised prostate cancer: the MRC PR06 trial. *BJU Int.* 94(9), 1400–1401 (2004).

29 Wallace K, Fleshner N, Jewett M, Basiuk J, Crook J. Impact of a multi-disciplinary patient education session on accrual to a difficult clinical trial: the Toronto experience with the surgical prostatectomy versus interstitial radiation intervention trial. *J. Clin. Oncol.* 24(25), 4158–4162 (2006).

30 Wilt TJ, Brawer MK, Barry MJ *et al.* The Prostate Cancer Intervention versus Observation Trial: VA/NCI/AHRQ

cooperative studies program #407 (PIVOT): design and baseline results of a randomized controlled trial comparing radical prostatectomy to watchful waiting for men with clinically localized prostate cancer. *Contemp. Clin. Trials* 30(1), 81–87 (2009).

31 Wilt TJ, Brawer MK, Jones KM *et al.* Radical prostatectomy versus observation for localized prostate cancer. *N. Engl. J. Med.* 367(3), 203–213 (2012).

■■ **One of two completed randomized trials of surgery versus observation for localized prostate cancer, demonstrating the difficulty in accruing for such trials and generalizing the results.**

32 Bill-Axelson A, Holmberg L, Ruutu M *et al.* Radical prostatectomy versus watchful waiting in early prostate cancer. *N. Engl. J. Med.* 364(18), 1708–1717 (2011).

■■ **One of two completed randomized trials of surgery versus observation for localized prostate cancer, demonstrating the difficulty in accruing for such trials and generalizing the results.**

33 Donovan J, Hamdy F, Neal D *et al.* Prostate testing for cancer and treatment (ProtecT) feasibility study. *Health Technol. Assess.* 7(14), 1–88 (2003).

34 Donovan J, Mills N, Smith M *et al.* Quality improvement report: improving design and conduct of randomised trials by embedding them in qualitative research: ProtecT (prostate testing for cancer and treatment) study. Commentary: presenting unbiased information to patients can be difficult. *BMJ* 325(7367), 766–770 (2002).

35 Lobo FS, Wagner S, Gross CR, Schommer JC. Addressing the issue of channeling bias in observational studies with propensity scores analysis. *Res. Social Adm. Pharm.* 2(1), 143–151 (2006).

36 Potosky AL, Harlan LC, Stanford JL *et al.* Prostate cancer practice patterns and quality of life: the prostate cancer outcomes study. *J. Natl Cancer Inst.* 91(20), 1719–1724 (1999).

37 Touijer K, Eastham JA, Secin FP *et al.* Comprehensive prospective comparative analysis of outcomes between open and laparoscopic radical prostatectomy conducted in 2003 to 2005. *J. Urol.* 179(5), 1811–1817; discussion 1817 (2008).

38 Kupelian PA, Potters L, Khuntia D *et al.* Radical prostatectomy, external beam radiotherapy <72 Gy, external beam radiotherapy > or =72 Gy, permanent seed implantation, or combined seeds/external beam radiotherapy for stage T1-T2 prostate cancer. *Int. J. Radiat. Oncol. Biol. Phys.* 58(1), 25–33 (2004).

39 Alemozaffar M, Regan MM, Cooperberg MR *et al.* Prediction of erectile function following treatment for prostate cancer. *JAMA* 306(11), 1205–1214 (2011).

■ **High-quality prospective cohort study with patient-reported outcomes in prostate cancer.**

40 Zhou P, Chen MH, McLeod D, Carroll PR, Moul JW, D'Amico AV. Predictors of prostate cancer-specific mortality after radical prostatectomy or radiation therapy. *J. Clin. Oncol.* 23(28), 6992–6998 (2005).

41 Wilt TJ, Shamliyan T, Taylor B *et al.* Comparative effectiveness of therapies for clinically localized prostate cancer. In: *Comparative Effectiveness Review*. Agency for Healthcare Research and Quality, Rockville, MD (2008).

42 Barocas DA, Penson DF. Racial variation in the pattern and quality of care for prostate cancer in the USA: mind the gap. *BJU Int.* 106(3), 322–328 (2010).

43 Porten SP, Cooperberg MR, Konety BR, Carroll PR. The example of CaPSURE: lessons learned from a national disease registry. *World J. Urol.* 29(3), 265–271 (2011).

44 Stier DM, Greenfield S, Lubeck DP *et al.* Quantifying comorbidity in a disease-specific cohort: adaptation of the total illness burden index to prostate cancer. *Urology* 54(3), 424–429 (1999).

45 Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med. Care* 30(6), 473–483 (1992).

46 McHorney CA, Ware JE, Raczek AE. The MOS 36-item short-form health survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Med. Care* 31(3), 247–263 (1993).

47 McHorney CA, Ware JE, Lu JF, Sherbourne CD. The MOS 36-item short-form health survey (SF-36): III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups. *Med. Care* 32(1), 40–66 (1994).

48 Szymanski KM, Wei JT, Dunn RL, Sanda MG. Development and validation of an abbreviated version of the expanded prostate cancer index composite instrument for measuring health-related quality of life among prostate cancer survivors. *Urology* 76(5), 1245–1250 (2010).

49 Kaplan SH, Greenfield S, Gandek B, Rogers WH, Ware JE. Characteristics of physicians with participatory decision-making styles. *Ann. Intern. Med.* 124(5), 497–504 (1996).

50 Haley SM, McHorney CA, Ware JE. Evaluation of the MOS SF-36 physical functioning scale (PF-10): I. Unidimensionality and reproducibility of the Rasch item scale. *J. Clin. Epidemiol.* 47(6), 671–684 (1994).

51 Roth AJ, Rosenfeld B, Kornblith AB *et al.* The memorial anxiety scale for prostate cancer. *Cancer* 97(11), 2910–2918 (2003).

52 Spencer BA, Steinberg M, Malin J, Adams J, Litwin MS. Quality-of-care indicators for early-stage prostate cancer. *J. Clin. Oncol.* 21(10), 1928–1936 (2003).

53 Penson DF. Assessing the quality of prostate cancer care. *Curr. Opin. Urol.* 18(3), 297–302 (2008).

54 Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap) – a metadata-driven methodology and workflow process for providing translational research informatics support. *J. Biomed. Inform.* 42(2), 377–381 (2009).

55 Li J, Djenaba JA, Soman A, Rim SH, Master VA. Recent trends in prostate cancer incidence by age, cancer stage, and grade, the United States, 2001–2007. *Prostate Cancer* 2012, 691380 (2012).

56 Ayanian JZ, Chrischilles EA, Fletcher RH *et al.* Understanding cancer treatment and outcomes: the cancer care outcomes research and surveillance consortium. *J. Clin. Oncol.* 22(15), 2992–2996 (2004).

57 Brawley OW. Trends in prostate cancer in the United States. *J. Natl Cancer Inst. Monogr.* 2012(45), 152–156 (2012).

58 Horn SD, Gassaway J. Practice based evidence: incorporating clinical heterogeneity and patient-reported outcomes for comparative effectiveness research. *Med. Care* 48(Suppl. 6), S17–S22 (2010).

59 Kaplan SH, Billimek J, Sorkin DH, Ngo-Metzger Q, Greenfield S. Who can respond to treatment? Identifying patient characteristics related to heterogeneity of treatment effects. *Med. Care* 48(Suppl. 6), S9–S16 (2010).

60 Miller DC, Spencer BA, Shah RB *et al.* The quality of surgical pathology care for men undergoing radical prostatectomy in the U.S. *Cancer* 109(12), 2445–2453 (2007).

61 Spencer BA, Miller DC, Litwin MS *et al.* Variations in quality of care for men with early-stage prostate cancer. *J. Clin. Oncol.* 26(22), 3735–3742 (2008).

62 Kaplan SH, Gandek B, Greenfield S, Rogers W, Ware JE. Patient and visit characteristics related to physicians' participatory decision-making style. Results from the medical outcomes study. *Med. Care* 33(12), 1176–1187 (1995).

63 Kaplan SH, Dukes KA, Sullivan LM, Tripp TJ, Greenfield S. Is passivity a risk factor for poor health outcomes? *J. Gen. Int. Med.* 11(Suppl. 1), 76 (1996).

■ Websites

101 Clinicaltrials.gov. http://clinicaltrials.gov

102 National Cancer Institute. Surveillance Epidemiology and End Results. http://seer.cancer.gov/registries

103 Howlader N, Noone AM, Krapcho M *et al.* (Eds). SEER Cancer Statistics Review, 1975–2010, National Cancer Institute. Bethesda, MD, USA. http://seer.cancer.gov/csr/1975_2010

104 United States Census Bureau. 2010 United States Census. http://quickfacts.census.gov/qfd/states/00000.html

105 Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services. http://aspe.hhs.gov/health/reports/2011/cpshealthins2011/ib.shtml