The Effectiveness of Screening for Prostate Cancer

A Nested Case-Control Study

John Concato, MD, MPH; Carolyn K. Wells, MPH; Ralph I. Horwitz, MD; David Penson, MD; Graeme Fincke, MD; Dan R. Berlowitz, MD, MPH; Gregory Froehlich, MD; Dawna Blake, MD; Martyn A. Vickers, MD; Gerald A. Gehr, MD; Nabil H. Raheb, MD; Gail Sullivan, MD, MPH; Peter Peduzzi, PhD

Background: Screening for prostate cancer is done commonly in clinical practice, using prostate-specific antigen (PSA) tests or digital rectal examination (DRE). Evidence is lacking, however, to confirm a survival benefit among screened patients. We evaluated the effectiveness of PSA, with or without DRE, in reducing mortality.

Methods: We conducted a multicenter nested case-control study at 10 Veterans Affairs medical centers in New England. Among 71,661 patients receiving ambulatory care between 1989 and 1990, 501 case patients were identified as men who were diagnosed as having adenocarcinoma of the prostate from 1991 through 1995 and who died sometime between 1991 and 1999. Control patients were men who were alive at the time the corresponding case patient had died, matched (1:1 ratio) for age and Veterans Affairs facility. The exposure variable (determined blind to case-control status) was whether PSA testing or DRE was performed for screening prior to the diagnosis of prostate cancer among case patients, with the same time interval for control patients. The association of screening and overall or cause-specific (prostate cancer) mortality was adjusted for race and comorbidity.

Results: A benefit of screening was not found in our primary analysis assessing PSA screening and all-cause mortality (adjusted odds ratio, 1.08; 95% confidence interval, 0.71-1.64; \( P = .72 \)), nor in a secondary analysis of PSA and/or DRE screening and cause-specific mortality (adjusted odds ratio, 1.13; 95% confidence interval, 0.63-2.06; \( P = .68 \)).

Conclusions: These results do not suggest that screening with PSA or DRE is effective in reducing mortality. Recommendations for obtaining “verbal informed consent” from men regarding such screening should continue.

Arch Intern Med. 2006;166:38-43

Among all types of cancer affecting American men in 2005, adenocarcinoma of the prostate gland is projected to rank first in incidence (232,090 diagnoses) and second in mortality (30,350 deaths). Measurement of prostate-specific antigen (PSA) in serum and digital rectal examination (DRE) are commonly used to screen for prostate cancer, yet official recommendations regarding these tests vary. For example, American Cancer Society recommendations include screening for prostate cancer in men older than 50 years, using PSA testing and DRE, followed by transrectal ultrasound if either test result is abnormal. In contrast, the American College of Physicians suggests counseling regarding possible benefits and risks, and the US Preventive Services Task Force found insufficient evidence to recommend screening. These positions were promulgated in the setting of data showing that the screening tests increase detection of prostate cancer but without direct evidence showing that PSA or DRE reduce mortality.

For editorial comment see page 7

Screening tests almost always increase the detection of disease, but among other requirements for improving survival among patients with cancer, the tumors detected must be both potentially fatal as well as curable. The lack of suitable evidence, from either randomized trials or well-designed observational studies, creates uncertainty regarding the effectiveness of screening for prostate cancer. The purpose of the current nested case-control study was to address the question: Does screening with PSA, with or without DRE, improve survival in clinical practice?

Author Affiliations are listed at the end of this article.
GENERAL STRATEGY

In a case-control analysis of the impact of screening on survival, previous exposure to screening tests is compared among people with cancer who died (case patients) vs a representative selection of people (control patients) who are still alive with or without the same cancer. A lower frequency, or deficit, of screening among case patients (vs controls) provides evidence supporting a protective effect of screening on mortality. The design of the present study has been reported previously. The Human Subjects Subcommittee at each institution approved the study protocol.

SAMPLING FRAME

The source population, identified using the Department of Veterans Affairs (VA) Outpatient Clinic System database, included men receiving ambulatory care between 1989 and 1990 at any of the 10 VA medical centers in New England (participating facilities are listed in a box on page 42). Patients were eligible for inclusion in the study cohort if they were at least 50 years old and without a diagnosis of prostate cancer as of January 1, 1991. These criteria identified a baseline population of men who were “candidates” for prostate cancer screening.

CASE AND CONTROL PATIENTS

Potential case patients were men in the cohort who had incident diagnoses of prostate cancer between January 1, 1991, and December 31, 1995, identified using pathology databases at each VA medical center. Among these men, cases were identified as those who died at any time during follow-up from diagnosis through December 31, 1999.

Control patients were selected at random from the remaining men in the cohort still receiving VA care, matched 1:1 for each case patient based on VA facility and age. Men with prostate cancer were eligible to be selected as controls, using the criterion that control patients had to be alive on the date of death of the corresponding matched case patient. If a control patient with prostate cancer later died, he was resampled as a case patient so that the impact of prostate cancer screening on mortality was determined in an unbiased manner.

A medical record review ascertained race and baseline comorbidity, as well as anatomic stage and histological grade of prostate cancer (for case patients). The coding of comorbidity involved assigning designations based on the Charlson index coded in 4 categories; a check of interobserver variability (for 49 records, or approximately 5% of men) for comorbidity found 86% agreement with a weighted k value of 0.80.

EVALUATION OF POTENTIAL SCREENING TESTS

Medical records (outpatient and inpatient) were reviewed thoroughly to determine whether patients had received screening with PSA or DRE. For each matched case-control pair, the earliest date for potential screening was January 1, 1991, for dates of diagnosis in 1992 through 1995; this start date was moved to July 1, 1990, for diagnoses in 1991. The latest date for potential screening was when a diagnosis of prostate cancer was made for each case patient, and the same date was used for his matched control. Thus, the interval of possible screening ranged from 6 months to 5 years. The requirement to link the interval for screening among control patients to that of each corresponding case patient protects against bias that would be introduced if screening were examined over a longer period for controls than for cases.

Information about screening was obtained by trained and experienced research associates who were “blind” to case or control status. The performance and results of PSA testing were verified using each VA facility’s laboratory database. Validated strategies were used to identify true screening tests from among all documented PSAs and DREs. For example, the date of suspicion of prostate cancer (that could precede, but not follow, the date of diagnosis) was established for all case patients, beyond which tests could not be classified as screening. The basis for this determination was evidence that the suspicion of cancer had increased above a baseline level. Any PSA test or DRE that led to a “surprise” diagnosis of cancer was considered screening.

CLASSIFICATION OF PATIENTS AS SCREENED OR NOT SCREENED

For potential screening tests, previous work had developed a mutually exclusive and exhaustive classification scheme for identifying screening tests in medical record reviews of patients with prostate cancer. The categories include definite screening; probable screening; tests done for differential diagnosis (nonscreening); and tests done for unknown reasons. Each patient was subsequently coded as screened vs not screened. Using approximately 3% of records (n=46) to assess interobserver variability for the process of classifying screening, we found 98% agreement and a weighted k value of 0.90.

DETERMINATION OF MORTALITY

A 4- to 9-year period of follow-up, ending in December 1999, was used to assess overall mortality as the primary outcome. The VA Beneficiary Identifier Locator System and Choice-Point (Alpharetta, Ga) were used to determine the vital status of patients (last queries done in 2002), without the need for continuous follow-up in the medical record. Data from these sources are more than 93% accurate.

In a secondary analysis, cause of death was assessed using a standardized extraction form and uniform coding criteria. Paper and electronic medical records were reviewed for evidence of active and progressive prostate cancer, using symptoms (eg, bone pain), imaging studies (eg, bone scan), laboratory results (eg, PSA and alkaline phosphatase), and documented impressions of clinicians (eg, whether the patient’s decline or death was attributed to prostate cancer). The same approach was used to ascribe death due to other causes. Consensus was reached for each patient’s medical record; categories of uncertain and unknown were also used.

STATISTICAL ANALYSES

The primary analysis examined “definite” PSA screening tests and overall mortality. We assumed an expected screening frequency of 30% among controls and designated an odds ratio (OR) of 0.67 as an important protective effect of screening, with a P = .05 (2 tailed) and 80% power. Five-year (overall) survival for men with prostate cancer was estimated as 65%; standard methods for calculating sample size for matched case-control studies determined that 498 cases and 498 controls were required. The data were analyzed using conditional logistic regression for matched data, adjusting for comorbidity and race. A secondary analysis addressed cause-specific (prostate cancer) mortality as the outcome and a broadened definition of screening to include PSA or DRE as a program of screening. The ORs from the primary and secondary analyses define a range of point estimates for the effectiveness of screening. Additional analyses involved screening redefined to include patients with symptoms of benign prostatic hyperplasia (a common clinical scenario); and subsets of case patients and matched
controls including those receiving a diagnosis between 1993 and 1995 (when screening was more common), those without substantial baseline comorbidity (a "sensitivity analysis" for cause-specific death), and younger patients. Finally, exposure was redefined as tests done for differential diagnosis, to check for an increased association with mortality.

**RESULTS**

**SOURCE POPULATION**

A total of 72,909 men who were born before or during 1940 had ambulatory care clinic visits at any of the 10 VA medical centers in New England between 1989 and 1990. Among this population, 1248 men were removed from the study population owing to a diagnosis of prostate cancer before 1991. The final cohort included the remaining 71,661 men, with 14,430 (20.1%) aged 50 to 59 years; 33,360 (46.6%), 60 to 69 years; 20,640 (28.8%), 70 to 79 years; and 3231 (4.5%), 80 years or older.

Among this group, 1425 men were diagnosed as having prostate cancer between 1991 and 1995. Follow-up for mortality through 1999 identified 501 case patients who died, exceeding our sample size threshold of 498. As of January 1, 2000, 5-year overall survival among the 1425 men with prostate cancer was 68%, with a projected median survival of 8.1 years (95% confidence interval [CI], 7.6-9.0 years). For each case patient, a control patient (n=501) matched for age and site was randomly selected from the cohort.

**PATIENT CHARACTERISTICS**

Demographic and clinical characteristics for the 1002 patients in case-control analyses are given in Table 1; the median age was 72.5 years. Cases, compared with controls including those receiving a diagnosis between 1993 and 1995 (when screening was more common), those without substantial baseline comorbidity (a “sensitivity analysis” for cause-specific death), and younger patients. Finally, exposure was redefined as tests done for differential diagnosis, to check for an increased association with mortality.

**SOURCE POPULATION**

A total of 72,909 men who were born before or during 1940 had ambulatory care clinic visits at any of the 10 VA medical centers in New England between 1989 and 1990. Among this population, 1248 men were removed from the study population owing to a diagnosis of prostate cancer before 1991. The final cohort included the remaining 71,661 men, with 14,430 (20.1%) aged 50 to 59 years; 33,360 (46.6%), 60 to 69 years; 20,640 (28.8%), 70 to 79 years; and 3231 (4.5%), 80 years or older.

Among this group, 1425 men were diagnosed as having prostate cancer between 1991 and 1995. Follow-up for mortality through 1999 identified 501 case patients who died, exceeding our sample size threshold of 498. As of January 1, 2000, 5-year overall survival among the 1425 men with prostate cancer was 68%, with a projected median survival of 8.1 years (95% confidence interval [CI], 7.6-9.0 years). For each case patient, a control patient (n=501) matched for age and site was randomly selected from the cohort.

**SECONDARY ANALYSES**

A secondary analysis examined the association of screening and cause-specific mortality (Table 2). Overall, 136 (27%) of 501 case patients had evidence of death due to prostate cancer. The adjusted OR for screening with PSA or DRE was 1.13 (95% CI, 0.63-2.06; P=.68). Black race...
(OR, 4.46; 95% CI, 1.39-14.3; P=.01) and comorbidity (OR, 1.26; 95% CI, 1.01-1.57; P=.04) remained statistically significant predictors of (cause-specific) mortality. Only 11% of deaths were classified as due to “unknown” causes, attributable mostly to unavailable medical records near the time of death. Results were similar (data not shown) when an analysis was done recoding unknown cause of deaths as related to prostate cancer.

To explore our data further, additional analyses were conducted (Table 3). When men with symptoms of benign prostatic hyperplasia were included in the screening group, the adjusted OR for screening with PSA and mortality was 1.22 (95% CI, 0.84-1.78; P=.30). Among the subset of case patients diagnosed with prostate cancer between 1993 and 1995 and their matched control patients, when PSA screening was done more frequently, the adjusted OR for the association of screening and mortality was 1.06 (95% CI, 0.66-1.72; P=.81). In the subset of case patients with no or mild comorbidity and their matched controls, the OR for the association of PSA screening and mortality was 1.12 (95% CI, 0.56-2.27; P=.75). In the subset of case patients aged 72.5 years or younger and their matched controls, the OR for the association of PSA screening and mortality was 1.11 (95% CI, 0.62-2.00; P=.73).

**Mortality Associated with Nonscreening Tests**

The adjusted OR for the association of tests done for differential diagnosis and mortality was 1.86 (95% CI, 1.18-2.93; P=.008). This analysis examines a plausible biological “mechanism”—patients who receive PSA or DRE tests when presenting with symptoms highly suspicious for cancer (rather than presenting for screening) should be, and were, more likely to die during a period of follow-up.

**Randomized Controlled Trials**

Randomized controlled trials are considered the “gold standard” research design, but reports suggest that well-designed observational studies can provide equally valid results. When results of observational studies are discordant with those from randomized trials, the non-random designs are generally considered prone to exaggerating (rather than obscuring) the benefit of an intervention. In contrast, the current nested case-control study of screening for prostate cancer found no evidence of a survival benefit associated with PSA testing or DRE. Importantly, similar case-control analyses have been endorsed when finding evidence in support of other screening tests, including Papanicolaou smears, mammograms, and sigmoidoscopies.

The only completed randomized trial on the topic of screening for prostate cancer had substantial design flaws. A case-control study found that metastatic prostate cancer was not reduced among men screened with DRE, but survival was not assessed. Another case-control study concluded that DREs were associated with reduced mortality, but an arbitrary cutoff 1 year before diagnosis was used to exclude diagnostic (vs screening) DREs, making the study vulnerable to misclassification of exposure to screening and mortality was 1.11 (95% CI, 0.62-2.00; P=.73).

**Table 2. Conditional Logistic Regression Models of Mortality**

<table>
<thead>
<tr>
<th>Mortality</th>
<th>No. (%), Factor In</th>
<th>Matched Adjusted OR* (95% CI)</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall mortality (501 pairs)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Screening with PSA</td>
<td>70 (14)</td>
<td>65 (13)</td>
<td>1.08 (0.71-1.64)</td>
</tr>
<tr>
<td>Black race</td>
<td>48 (10)</td>
<td>21 (4)</td>
<td>3.18 (1.74-5.83)</td>
</tr>
<tr>
<td>Moderate to severe comorbidity</td>
<td>240 (48)</td>
<td>144 (29)</td>
<td>1.46 (1.31-1.64)</td>
</tr>
<tr>
<td>Cause-specific mortality (136 pairs)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Screening with PSA/DRE</td>
<td>61 (45)</td>
<td>56 (41)</td>
<td>1.13 (0.63-2.06)</td>
</tr>
<tr>
<td>Black race</td>
<td>15 (11)</td>
<td>5 (4)</td>
<td>4.46 (1.39-14.3)</td>
</tr>
<tr>
<td>Moderate to severe comorbidity</td>
<td>52 (38)</td>
<td>38 (28)</td>
<td>1.26 (1.01-1.57)</td>
</tr>
</tbody>
</table>

Abbreviations: CI, confidence interval; DRE, digital rectal examination; OR, odds ratio; PSA, prostate-specific antigen.

*Case and control patients were matched for age and facility; results were adjusted for variables in table.

**Table 3. Additional Analyses of Screening and Overall Mortality in Selected Groups**

<table>
<thead>
<tr>
<th>Group</th>
<th>Screening Frequency, No. (%)</th>
<th>Matched Adjusted OR* (95% CI)</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Screening includes BPH (n = 1002)</td>
<td>102 (20)</td>
<td>87 (17)</td>
<td>1.22 (0.84-1.78)</td>
</tr>
<tr>
<td>Diagnosis in 1993-1995 (n = 508)</td>
<td>53 (21)</td>
<td>53 (21)</td>
<td>1.06 (0.66-1.72)</td>
</tr>
<tr>
<td>Comorbidity none/mild (n = 522)</td>
<td>33 (13)</td>
<td>33 (13)</td>
<td>1.12 (0.56-2.27)</td>
</tr>
<tr>
<td>Age &lt;72.5 y (n = 502)</td>
<td>40 (16)</td>
<td>33 (13)</td>
<td>1.11 (0.62-2.00)</td>
</tr>
</tbody>
</table>

Abbreviations: BPH, benign prostatic hyperplasia; CI, confidence interval; OR, odds ratio.

*Odds ratio for screening; matched for age and facility, and adjusted for race and comorbidity.

Findings from a third study examining DRE screening in the pre-PSA era were also negative. A more recent case-control study of DRE and PSA found a protective benefit of overall screening, but the result did not achieve statistical significance. PSA screening was uncommon, and comorbidity was not examined thoroughly.

The present investigation of screening for prostate cancer used a case-control design nested within a cohort of patients receiving medical care in a large geographic region. Several strategies, organized in categories, were adopted to avoid methodological bias. The first category involved baseline characteristics—matching for age and adjusting for differences in race and comorbidity. These approaches account for factors related to the likelihood of receiving screening tests as well as to the risk of dying (ie, confounders).

The second category was related to exposure to screening, and involved enumerating individual screening tests as well as a patient-specific classification of exposure to screening.
screening. Prior studies\textsuperscript{38,20} have often used arbitrary cut-off dates to define screening—an approach that is prone to misclassification. Our review of medical records involved an evaluation of the clinical circumstances of each test (ie, clinic location, results of prior PSA or DRE tests, and assessment of clinician). In addition, a date of suspicion for each patient with prostate cancer was determined, beyond which tests could not be classified as screening. This approach addresses concerns regarding the detectable curable preclinical phase of a cancer.\textsuperscript{33} Importantly, restrictive definitions of screening—as in our primary analysis—should bias results in favor of screening.\textsuperscript{34} We also conducted analyses of PSA screening alone as well as PSA or DRE screening (representing the patterns of screening most likely to be used in clinical practice), and we varied the definition of screening based on whether patients had lower urinary tract symptoms; our results were similar in all scenarios.

The third category involved overall mortality as the primary outcome. In particular, assembly of a source population (cohort) protects against lead time and length bias\textsuperscript{35} by assessing men’s longevity rather than diagnosis-to-death intervals. (Lead-time bias refers to screening extending a man’s diagnosis-to-death interval, without improving actual longevity. Length bias refers to screening detecting slower growing tumors that are less likely to be fatal.) Although overall mortality is not always used in studies of screening, death per se is most relevant to patients, and comorbidity is a strong determinant of survival among men with prostate cancer.\textsuperscript{36} In addition, although death certificates can be used to attribute cause of death,\textsuperscript{37} opportunities for differential error exist.\textsuperscript{38} We therefore used a medical record review to assess the impact of screening on cause-specific mortality (Table 2). A related secondary analysis (Table 3) focusing on patients with limited or no comorbidity (representing the group most likely to experience mortality due to prostate cancer) provides important confirmation of this finding.

Our analysis is also unlikely to be influenced by several other sources of bias. For example, ascertainment bias\textsuperscript{39} is proposed to occur if screening occurs before, or cancer is diagnosed after, the period of enrollment used in the study. Prostate-specific antigen, however, was used infrequently as a screening test prior to our study’s “starting date” in 1991. In addition, we required controls to be alive at the time of the corresponding case patient’s death so that men who died between 1995 and 1999 were eligible to be selected as controls, as suggested.\textsuperscript{39}

The decision to match on facility might be viewed as a limitation, but this approach was used to account for clustering of patients by site. In addition, although the proportion of “screened patients” in our study (Figure) was relatively low, the frequency of patients receiving PSA (and DRE) tests was higher (data not shown), and analyses of Medicare data\textsuperscript{40} found rates of testing similar to the current analyses. (Of note, VA guidelines promote counseling regarding screening rather than routine testing.) Even when the overall prevalence of screening in our study exceeded 40%, however, a benefit of screening was still not evident (Table 2). Finally, patients for this study were restricted to men who received ambulatory care in the VA system. Although less than 1% of men had progress notes indicating that PSA tests were done in other clinical settings, data for non-VA screening were not available for analysis.

The 4- to 9-year period for mortality ascertainment is consistent with the study design because finding unscreened men who died relatively soon after diagnosis is needed to establish a benefit of screening. The fact that a similar (rather than a higher) proportion of control patients, selected randomly from the source population, received screening is the crux of our “negative” study findings (Figure). The lower bounds of CIs for screening include a protective effect (Table 2 and Table 3), but ORs consistently around the null value do not suggest a benefit of PSA testing or DRE. Results linking tests done for “diagnostic” diagnosis with increased mortality support the validity of our analyses.

Innovations in PSA testing since the early 1990s, such as PSA density or velocity and free vs bound PSA, can increase the sensitivity or specificity of detecting prostate cancer, but evidence is not available currently to link these techniques with improved survival. Ecological investigations\textsuperscript{41,42} of population level rates of screening for prostate cancer and disease-specific mortality have provided conflicting results, but such studies are much more vulnerable to bias compared with observational studies examining cause-effect associations in individual patients. Other case-control studies and 2 randomized trials\textsuperscript{43,44} are in progress to assess the impact of PSA testing or DRE on survival. Whether the case-control studies show a benefit of screening, strengths and limitations of observational studies will be pertinent issues for discussion. Similarly, strengths (eg, randomization) and limitations (eg, contamination of screening in the no-screening arm) of ongoing clinical trials are also relevant.

Optimal clinical strategies for diagnosing and treating prostate cancer remain uncertain and in need of additional investigation. Based on available evidence, including the present study, recommendations regarding screening for prostate cancer should not endorse routine testing of asymptomatic men to reduce mortality. Rather, the uncertainty of screening should be explained to patients in a process of “verbal informed consent,” promoting informed decision making.\textsuperscript{45}

\textsuperscript{Connecticut:} Newington and West Haven; \textsuperscript{Maine:} Togus; \textsuperscript{Massachusetts:} Bedford, Boston, Brockton/West Roxbury, and Northampton; \textsuperscript{New Hampshire:} Manchester; \textsuperscript{Rhode Island:} Providence; and \textsuperscript{Vermont:} White River Junction.
Accepted for Publication: August 4, 2005.

Author Affiliations: Clinical Epidemiology Research Center and Cooperative Studies Program Coordinating Center, Department of Veterans Affairs Connecticut Healthcare System, West Haven Veterans Affairs Medical Center (VAMC), West Haven (Drs Condatco and Peduzzi and Ms Wells); Department of Medicine, Yale University School of Medicine, New Haven, Conn (Drs Condatco and Peduzzi and Ms Wells); Office of the Dean, Case Western Reserve School of Medicine, Cleveland, Ohio (Dr Horwitz); Section of Urology, University of Southern California Keck School of Medicine, Los Angeles (Dr Penson); Center for Health Quality, Outcomes, and Economic Research, Bedford, Mass (Drs Fincke and Belowitz); Veterans Affairs Outcomes Research Group, White River Junction, Vt (Dr Froehlich); Providence Veterans Affairs Medical Center, Providence, RI (Dr Blake); Surgical Service, Togus VAMC, Togus, Me (Dr Vickers); Medical Service, Manchester VAMC, Manchester, NH (Dr Gehr); Surgical Service, Northwestern VAMC, Lees, Mass (Dr Raheb); VA Connecticut Healthcare System, Newington VAMC, Newington, Conn (Dr Sullivan).

Correspondence: John Condatco, MD, MPH, VA Connecticut Healthcare System, Clinical Epidemiology Research Center (Mail Code 151B), 950 Campbell Ave, West Haven, CT 06516 (john.condatco@yale.edu).

Financial Disclosure: None.

Funding/Sponsorship: This research was supported by grant funding from the Department of Veterans Affairs.

Acknowledgment: We wish to acknowledge invaluable contributions from K. Anderson, G. Castellazzo, D. Cavaliere, P. Crutchfield, N. Cummings, J. Fino, A. Kamina, V. Latvis, G. McaVay, J. Mezger, D. Orlando, R. Roy, J. Talarczyk, L. Thomas, C. Ufferliffe (deceased), and I. Vonnick at West Haven, Conn; E. Beaupre at Togus, Me; J. Jacobson and S. Kumar at West Roxbury, Mass; R. Pritchard at White River Junction, Vt; M. Ratheith at Newington, Conn; and 4 anonymous reviewers.

REFERENCES


