Tuukka Mäkinen,2 Teuvo L. J. Tammela, Matti Hakama, Ulf-Håkan Stenman, Sakari Rannikko, Jussi Aro, Harri Juusela, Liisa Määttänen, and Anssi Auvinen

Department of Surgery, Seinäjoki Central Hospital, FIN-60220 Seinäjoki [T. M.]; Department of Urology, Tampere University Hospital and Medical School [T. L. J. T.] and Tampere School of Public Health [M. H., A. A.], University of Tampere, Tampere; Departments of Clinical Chemistry [U-H. S.] and Urology [S. R., J. A.], Helsinki University Central Hospital, Helsinki; Department of Surgery, Jorvi Hospital, Helsinki University Central Hospital, Espoo [H. J.]; and Finnish Cancer Registry, Helsinki [L. M.], Finland

ABSTRACT

Purpose: Early diagnosis of prostate cancer is a necessary, but not sufficient, prerequisite for an effective screening program aiming at mortality reduction. We compared tumor characteristics between the screening and control arms in the Finnish population-based screening trial.

Experimental Design: The Finnish trial is the largest component in the European Randomized Study of Screening for Prostate Cancer. A total of 24,000 men aged 55–67 years were randomized to the screening arm, whereas 35,973 men formed the control arm during the first three screening years. At the time of invitation, 22,732 men were eligible for screening, and 15,685 (69%) participated. A prostate-specific antigen (PSA) concentration of ≥4 μg/liter was defined as a screening-positive finding.

Results: The detection rate among screenees was 2.4% (377 of 15,685), whereas 0.6% (40 of 7,047) of nonparticipants in the screening arm and 0.3% (112 of 35,973) of the controls were diagnosed with prostate cancer during the first postrandomization year in the absence of screening. In the screening arm, 82% of the cancers were clinically organ confined compared with 65% in the control arm. Yet, both the absolute number and cumulative incidence of advanced cancer were higher in the screening arm. No differences were seen in the WHO grade distribution between the study groups. The median PSA was substantially lower among screen-detected cases (7.1 μg/liter) than among nonattenders (15.7 μg/liter) and controls (13.2 μg/liter).

Conclusions: Our findings on intermediate indicators of PSA screening provide encouraging, yet inconclusive evidence for eventual mortality reduction.

INTRODUCTION

The aim of prostate cancer screening is to reduce mortality through curative treatment of early stage disease. A decrease in prostate cancer mortality has been reported from the United States after adoption of widespread PSA testing (1). However, similar changes have been reported from other countries with less aggressive use of PSA, which suggests that mortality reduction may be also attributable to other factors, such as concurrent changes in treatment, e.g., early endocrine treatment (2). No mortality results are available from randomized trials, apart from early and inconclusive findings from the Quebec trial (3).

The Finnish population-based screening trial, with a total sample size of ~80,000 men, forms the largest component of the European Randomized Study of Screening for Prostate Cancer (4). Although the importance of randomized PSA screening trials has been well recognized, the present study provides to our knowledge the first intention-to-screen analysis of tumor characteristics in a prostate cancer screening trial.

MATERIALS AND METHODS

Study Population. The Finnish component of the European Randomized Study of Screening for Prostate Cancer was started in 1996. A total of 60,211 men aged 55–67 years (born 1929–1944) were enrolled from the Finnish Population Register in the period between 1996 and 1998. Men with prevalent prostate cancer (n = 238) were identified through record linkage with the Finnish Cancer Registry and excluded from the study before randomization. Annually, 8,000 men aged 55, 59, 63, and 67 years were randomized to the screening arm, and the remaining ~12,000 men formed the control arm. Men deceased, moved outside the study area by the time of invitation, or refusing the use of their address for any purpose were considered ineligible and not invited for screening (n = 1,268). Of the 22,732 men eligible for screening, 15,685 (69%) eventually participated. The 35,973 men comprising the control arm of the trial were not contacted (Fig. 1).

Information on prostate cancer in the control population and among nonparticipants was obtained through a record linkage with the Finnish Cancer Registry, which is a nationwide population-based cancer registry with virtually complete cover-

1 Supported by the Academy of Finland, the Cancer Society of Finland, the Medical Research Fund of Tampere University Hospital, the Europe Against Cancer Program, Beckman-Hybritech Corp., and the Sigrid Juselius Foundation. Dr. Mäkinen received additional support from the Cancer Society of Pirkanmaa and Medical Research Fund of Seinäjoki Central Hospital.

2 To whom requests for reprints should be addressed, at Seinäjoki Central Hospital, Hanneksenrinne 7, FIN-60220 Seinäjoki, Finland. Phone: 358-6-415 4111; Fax: 358-6-415 4040; E-mail: tuukka.makinen@epsh.fi.

3 The abbreviations used are: PSA, prostate-specific antigen; DRE, digital rectal examination; TNM, Tumor-Node-Metastasis; TRUS, transrectal ultrasound.
age of solid cancer cases in Finland (5). Information on cases among the screening participants was collected prospectively at participating hospitals. To ensure completeness of information, a record linkage with the discharge database of hospitals in the study area was conducted. The medical records were reviewed to obtain comparable information on stage and grade for patients detected outside the organized screening, i.e., in the control arm and among the nonparticipants. Causes leading to a diagnosis of prostate cancer were retrieved to assess the extent of PSA testing in the unscreened population. Opportunistic PSA screening was defined as a PSA determination in asymptomatic men during a general health checkup or at a physician’s appointment unrelated to any urological symptoms. One prostate cancer found at autopsy in the control arm was excluded.

**Screening Algorithm.** On informed consent, a blood sample was drawn from the screenees, and the serum PSA concentration was determined (Hybritech Tandem-E). All screening participants with a PSA of \( \geq 4 \) µg/liter were referred for diagnostic examinations, including DRE, TRUS, and prostate sextant biopsies. A directed biopsy was taken if a focal finding in either DRE or TRUS was noted. A supplemental DRE was offered for those with a PSA level of 3–3.9 µg/liter, and prostate biopsies were indicated if nodularity, induration, or asymmetry was present.

**Diagnostics.** All diagnoses were based on histological examination. Clinical staging at diagnosis was conducted according to the TNM classification primarily with TRUS and bone scan, but when necessary, other modalities were also used (6). The histological characteristics of detected tumors at biopsy were graded according to the WHO system (7).

**Statistics.** Screen-detected cancers were diagnosed in accordance with the study protocol within 12 months from drawing the blood sample. Among nonparticipants and controls, all prostate cancer cases detected during the first postrandomization year were included in the analyses. Clinical grade and stage of tumors detected in the screening and control arms were compared using Pearson’s \( \chi^2 \) test. Patients with unavailable clinical grade or stage were excluded from analyses. The proportions of organ-confined tumors and clinical grades were given with 95% confidence intervals. Cumulative incidence was defined as the

---

**Figure 1** Study protocol of the Finnish screening trial for prostate cancer from 1996 to 1998.
number of cases detected during the follow-up period (i.e., 12 months) relative to the number of men within a study group. The Wilcoxon signed rank test was used for comparison of PSA concentrations. Statistical analyses were performed on CIA version 1.1 (Martin J. Gardner and British Medical Journal) and S-PLUS version 4.0 (MathSoft, Inc., Cambridge, MA).

Ethics. The study protocol was approved by the ethical committee in each participating hospital. Permission to retrieve medical records was acquired from the Ministry of Social Affairs and Health and for cancer registry data from the Research and Development Center for Welfare and Health (STAKES).

RESULTS

A total of 377 prostate cancers were detected among the 15,685 screening participants, corresponding to a detection rate of 2.4%. Forty prostate cancers were diagnosed among the 7,047 nonparticipants (0.6%) in the screening arm and 112 cases among the 35,973 men (0.3%) in the control arm during the first postrandomization year.

More than half of the tumors outside the organized screening program were detected on the basis of lower urinary tract symptoms. Opportunistic PSA screening contributed to the diagnosis in 13% (5 of 40) of cases among the screening nonparticipants and in 21% (23 of 112) of the patients in the control arm.

Two-thirds of the screen-detected cases had a PSA level \(<10\, \mu g/liter\), as compared with one-fourth of the cases detected otherwise (Table 1). The median PSA was substantially lower among screen-detected cases (7.1 \(\mu g/liter\)) than among nonattenders (15.7 \(\mu g/liter, P < 0.001\)) and controls (13.2 \(\mu g/liter, P < 0.001\)). Overall, the difference between the screening and control arms was also substantial (medians 7.7 versus 13.2 \(\mu g/liter, P < 0.001\)).

Of the screen-detected prostate cancers, 85% (95% confidence interval 81–88%) were clinically organ confined, compared with 58% (41–73%, \(P < 0.001\)) among the nonattenders and 65% (57–74%, \(P < 0.001\)) in the control arm (Table 2). The overall proportion of organ-confined tumors in the screening arm was 82% (78–86%) based on intention-to-screen analysis (\(P < 0.001\)). Yet, both the absolute number (75 versus 38 cases) and number of nonlocal cancers relative to the number of men (cumulative incidence 0.3 versus 0.1%) were higher in the screening than control arm of the trial.

WHO grade I cancers comprised 37% (32–42%), grade II 55% (50–61%), and grade III 7% (5–11%) of the screen-detected tumors. The grade distribution of cases diagnosed among nonparticipants (\(P = 0.17\)) and controls (\(P = 0.71\)) was not different from that of the screen-detected tumors (Table 2). However, the number (32 versus 8 cases) as well as cumulative incidence (0.1 versus 0.02%) of poorly differentiated cancer were higher in the screening arm than among controls.

### Table 1

<table>
<thead>
<tr>
<th>PSA ((\mu g/liter))</th>
<th>Participants</th>
<th>tally</th>
<th>(%)</th>
<th>Nonparticipants</th>
<th>Control arm</th>
</tr>
</thead>
<tbody>
<tr>
<td>(&lt;4)</td>
<td>24 (6)</td>
<td>1 (3)</td>
<td>8 (7)</td>
<td>30 (27)</td>
<td></td>
</tr>
<tr>
<td>4–9.9</td>
<td>223 (59)</td>
<td>9 (23)</td>
<td>30 (75)</td>
<td>74 (66)</td>
<td></td>
</tr>
<tr>
<td>(\geq 10)</td>
<td>130 (34)</td>
<td>30 (75)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>377 (100)</td>
<td>40 (100)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*PC, prostate cancer.*

### Table 2

<table>
<thead>
<tr>
<th>Stage</th>
<th>Participants</th>
<th>Nonparticipants</th>
<th>Control arm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No. of PC (%)</td>
<td>No. of PC (%)</td>
<td>No. of PC (%)</td>
</tr>
<tr>
<td>T1N0M0</td>
<td>176 (47)</td>
<td>12 (30)</td>
<td>35 (31)</td>
</tr>
<tr>
<td>T2N0M0</td>
<td>143 (38)</td>
<td>11 (28)</td>
<td>37 (33)</td>
</tr>
<tr>
<td>T1–4N0M0</td>
<td>58 (15)</td>
<td>17 (43)</td>
<td>38 (34)</td>
</tr>
<tr>
<td>Unknown</td>
<td>2 (2)</td>
<td></td>
<td>2 (2)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Grade</th>
<th>Participants</th>
<th>Nonparticipants</th>
<th>Control arm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No. of PC (%)</td>
<td>No. of PC (%)</td>
<td>No. of PC (%)</td>
</tr>
<tr>
<td>I</td>
<td>139 (37)</td>
<td>20 (50)</td>
<td>45 (40)</td>
</tr>
<tr>
<td>II</td>
<td>209 (55)</td>
<td>16 (40)</td>
<td>56 (50)</td>
</tr>
<tr>
<td>III</td>
<td>28 (7)</td>
<td>4 (10)</td>
<td>8 (7)</td>
</tr>
<tr>
<td>Unknown</td>
<td>1 (0)</td>
<td></td>
<td>3 (3)</td>
</tr>
<tr>
<td>Total</td>
<td>377 (100)</td>
<td>40 (100)</td>
<td>112 (100)</td>
</tr>
</tbody>
</table>

*PC, prostate cancer.*
DISCUSSION

The aim of our study was to compare stage and grade of cancers detected in the screening and control arms of a randomized PSA-based screening trial. Previous studies on screening have reported a drift toward earlier stages using historical or otherwise selected hospital-referred patients as a control population (8, 9). However, analyses of time trends are not particularly informative in view of the lack of comparability (contemporaneous changes in factors other than screening, e.g., staging procedures or classifications), relatively low coverage of screening (reducing contrast and, hence, statistical power), as well as possible overdiagnosis (detection of indolent cases; Ref. 10). The same limitations also apply to geographical comparisons. Our results were obtained in a randomized, population-based screening trial. The participation rate was high, which is an essential requirement for a population-based (effectiveness) trial based on a study cohort representative of the general (source) population. The advantages of this experimental design include comparability of screening and control arms, i.e., balanced distribution of other factors ensured by randomization. This avoids the selection bias inherent in screening programs recruiting volunteers, e.g., participation affected by education, health insurance, and family history influencing the likelihood of prostate cancer diagnosis and death (11). Such selection was also evident in our results, as witnessed the large proportion of advanced cases among the nonparticipants. The advantages of randomization are lost if the analysis is not based on the intention-to-screen principle. An example of this is the Quebec trial, with mortality comparisons between screened and unscreened men irrespective of the result of the randomization (i.e., the trial arm to which they were allocated; Ref. 3).

Our results revealed a substantially smaller proportion of advanced prostate cancer in the screening than in the control arm (18 versus 35%) but no reduction in cumulative incidence of advanced prostate cancer. Although cancers in nonparticipants were more frequently advanced than screen-detected cases, this did not substantially affect the overall stage distribution in the screening arm because of their relatively small number. Stage of prostate cancer represents a surrogate measure of the effectiveness of screening, because curative treatment is available only for patients with organ-confined disease. Hence, a larger proportion of organ-confined cases is a prerequisite for effectiveness of screening. Although effective screening requires case detection at an earlier stage, a favorable shift in stage distribution is not sufficient evidence of mortality reduction. Screening is likely to cause lead-time bias because of the slow development of prostate cancer, i.e., only the survival time with disease is extended even if death is not postponed. Furthermore, detection of indolent cancers may artificially improve stage distribution and apparent survival. It also remains to be shown that rapidly growing aggressive cancers can be detected by screening at a curable stage. To reduce deaths from prostate cancer, screening will not only have to achieve detection at an early stage but also prevent deaths by altering the course of the disease.

A larger number and higher cumulative incidence of advanced cancer were seen in the screening than control arm, despite the fact that no information was available on interval cancers in the screened group. If screening succeeds in early detection and advancing the time of cancer diagnosis (as intended), it should be followed by a reduction in incidence in screened population. Therefore, the difference between the arms should diminish over the entire 4-year screening interval. We do not think that the lack of reduction in advanced cancer represents a failure of the screening, because of the fact that the lead time (i.e., advancement of diagnosis in time by screening) for clinically significant prostate cancer may be up to 10 years (12), allowing ample time for the control arm to catch up the difference in cumulative incidence. Rather, increased detection of advanced cancer may indicate that because of differences in natural course of the diseases, e.g., growth rate, the same process measures that are useful in breast cancer screening (13, 14) are probably not useful in prostate cancer screening.

How much can be inferred from these findings as to the effectiveness of screening also depends on the extent to which PSA, clinical stage, and grade predict mortality (predictive validity). They are the most powerful prognostic factors but, nevertheless, do not accurately predict the outcome (15). This limitation is most evident in clinically localized disease, because many cases are eventually upstaged and upgraded based on examinations of radical prostatectomy specimens (16). It is also unclear how applicable findings based on clinically detected cases are in the context of screen-detected cancer. A marked discrepancy has been observed between the prevalence of autopsy tumors and clinically detected cases, which suggests a strong possibility of overdiagnosis in screening (17). The detection of slow-growing, indolent tumors exposes the target population to unnecessary therapy and resultant morbidity and increases the costs of screening disproportionately. Previous screening studies have, however, suggested that the majority of screen-detected tumors are clinically significant in regard to tumor grade and stage (18). This notwithstanding, no method is currently available for reliable prediction of the significance of screen-detected tumors.

Information on screen-detected cancers was obtained prospectively, unlike those detected outside the organized screening program. However, a record linkage both with the Finnish Cancer Registry and discharge databases of hospitals in the study area ensured a high completeness of case ascertainment in both study arms. Yet, a higher completeness in the screening arm is possible, but it is unlikely to affect our conclusions unless very selective in terms of tumor characteristics. A limitation of our results is the fact that the cases did not undergo a central, blinded pathological evaluation. Yet, the same pathologists evaluated cancers in both arms using identical criteria. However, stage and grade were classified without blinding in regard to screening history. Hence, both misclassification and information bias are possible. With this in mind, we are planning a blinded central review of the histological specimens.

A potential source of bias in randomized screening trials is contamination, i.e., the use of PSA testing for opportunistic screening in the control population. Early detection with PSA has become a common practice especially in the United States, and hence, an unscreened control population is difficult to enroll for studies on prostate cancer screening. One of the strengths of our study is that PSA screening has been opposed as a public health policy in Finland. During the first three screening years,
only approximately a fifth of tumors in the control arm were attributable to contamination. This has to be taken into account when evaluating the long-term effects of screening.

Our results pertain to the first years of a screening program. Although the stage characteristics of tumors discovered in the control arm would not be expected to change with time, stage distribution among screen-detected tumors at subsequent screening rounds is likely to shift further to earlier stages (19, 20). Unless PSA-based case finding increases dramatically in the control arm, it is therefore likely that the difference in tumor stage between the screening and control arm will increase with time.

In conclusion, the Finnish screening trial with PSA provides encouraging evidence in terms of stage reduction, but definitive conclusions on the effectiveness of a screening program must be based on a comparison of prostate cancer mortality between screening and control arms during longtime follow-up.

REFERENCES
