Keywords:
Prostate cancer
Epidemiology
Prospective registries
Population-based
Disease-specific outcomes
Statistical approaches

Abstract

Context: Disease-specific registries that enroll a considerable number of patients play a major role in prostate cancer (PCa) research.

Objective: To evaluate available registries, describe their strengths and limitations, and discuss the potential future role of PCa registries in outcomes research.

Evidence acquisition: We performed a literature review of the Medline, Embase, and Web of Science databases. The search strategy included the terms prostate cancer, outcomes, statistical approaches, population-based cohorts, registries of outcomes, and epidemiological studies, alone or in combination. We limited our search to studies published between January 2005 and January 2015.

Evidence synthesis: Several population-based and prospective disease-specific registries are currently available for prostate cancer. Studies performed using these data sources provide important information on incidence and mortality, disease characteristics at presentation, risk factors, trends in utilization of health care services, disparities in access to treatment, quality of care, long-term oncologic and health-related quality of life outcomes, and costs associated with management of the disease. Although data from these registries have some limitations, statistical methods are available that can address certain biases and increase the internal and external validity of such analyses. In the future, improvements in data quality, collection of tissue samples, and the availability of data feedback to health care providers will increase the relevance of studies built on population-based and disease-specific registries.

Conclusions: The strengths and limitations of PCa registries should be carefully considered when planning studies using these databases. Although randomized controlled trials still provide the highest level of evidence, large registries play an important and growing role in advancing PCa research and care.

Patient summary: Several population-based and prospective disease-specific registries for prostate cancer are currently available. Analyses of data from these registries yield information that is clinically relevant for the management of patients with prostate cancer.

© 2015 European Association of Urology. Published by Elsevier B.V. All rights reserved.

* Corresponding author. Unit of Urology/Division of Oncology, San Raffaele Hospital, Milan, Italy; b Section of Cancer Surveillance, International Agency for Research on Cancer, Lyon, France; c Departments of Urology and Epidemiology & Biostatistics, Helen Diller Family Comprehensive Cancer Center, San Francisco, CA, USA; d Department of Urology, Mayo Clinic, Rochester, MN, USA; e Department of Urology, Queen’s University, Kingston, Ontario, Canada; f South Australian Prostate Cancer Clinical Outcomes Collaborative, Repatriation General Hospital, Daw Park, and the University of South Australia and the University of Adelaide, South Australia, Australia; g Division of Cancer Surgery, University of Melbourne, Peter MacCallum Cancer Centre, Melbourne, Australia; h Department of Urologic Surgery, Vanderbilt University, and the VA Tennessee Valley Geriatric Research, Education, and Clinical Center (GRECC), Nashville, TN, USA; i Division of Urologic Oncology, Department of Urology, University of Michigan, Ann Arbor, MI, USA
1. Introduction

Although several management options with improving long-term outcomes are currently available for patients with prostate cancer (PCa) [1], further research is still needed to improve the clinical management of this disease.

Historically, important research topics were investigated by studies that evaluated cohorts from single referral institutions. Numerous limitations often preclude generalization of results obtained in this setting. Similarly, the inclusion of highly selected patient groups, as well as costs and feasibility issues, can limit the validity of the small number of randomized controlled trials (RCTs) in the field of PCa [2]. The increasing availability of cancer registries, defined as organized systems that collect uniform data for a population defined by a particular disease, together with the improvements in data processing capabilities, has transformed PCa outcomes research during the last two decades [3–5].

The aim of this review is to evaluate currently available population-based and prospective disease-specific registries, to describe their strengths and limitations, to illustrate the types of studies that can be performed using these data, and to discuss the potential role of PCa registries in outcomes research in the future.

2. Evidence acquisition

A literature review was performed in January 2015 using the Medline, Embase, and Web of Science databases. The search strategy included the terms prostate cancer, outcomes, statistical approaches, population-based cohorts, registries of outcomes, and epidemiological studies, alone or in combination. We limited our search to population-based studies and investigations performed using prospective PCa registries published from January 2005 to January 2015. References cited in selected articles and in review articles retrieved in our search were also used to identify manuscripts that were not included in the initial search. The articles that provided the highest level of evidence were then evaluated and selected with the consensus of all authors of this manuscript. A total of 103 articles were reviewed.

3. Evidence synthesis

3.1. A role for PCa registries

According to current guidelines, the highest level of evidence and strongest grade of recommendation are provided by results of RCTs or meta-analyses of such studies [1]. Nonetheless, several issues often preclude generalization of results obtained in RCTs. First, these studies are in part limited by poor accrual; in fact, approximately 20% of adult cancer trials are never completed [6]. Second, patients participating in RCTs are often highly selected and might substantially differ from those seen in routine practice [3,7–11]. Third, RCTs are expensive. As a consequence, industry-funded studies are common in this setting [12]. However, sponsored trials are more likely to be published if positive in comparison to independent studies, which can be another source of bias [13,14]. Finally, RCTs, particularly in early PCa, take a long time to complete. Therefore, results from these studies might be obsolete by the time sufficient follow-up is achieved.

Observational studies represent an alternative to RCTs. Such studies are usually characterized by lower costs, higher patient numbers, more rapid accrual, and consequently a shorter time for identification and dissemination of results [15,16]. However, despite statistical controls, selection bias may affect results from single- and multi-institutional series [16,17]. In addition, most observational studies generally include men treated at high-volume tertiary referral centers. Since surgeon, radiotherapist, and oncology expertise, as well as hospital case volumes, affect treatment-related outcomes [18–21], results obtained in this setting might not be applicable to the general population. Unlike the majority of cancer data sets from large, highly specialized, single-center academic or tertiary referral institutions in the USA and Europe, registries reflect outcomes in men with PCa treated in real-world community settings. Moreover, because the data are primarily community- or population-based, they represent a meaningful standard of comparison for benchmarking at the individual, local, regional, or national level.

The significant practical limitations of RCTs and the bias and applicability concerns that may plague single-center cohort studies highlight the need for other sources of data to study PCa screening, diagnosis, treatment, and outcomes. PCa registries provide such an alternative. The democratization of patient management to a larger or general population creates a more generalizable pool for analysis and subsequent conclusions [3,4,15].

3.2. Types of PCa registries

PCa registries include both population-based and/or community cohorts and prospective patient registries. The term population-based refers to the systematic and ongoing collection of data on all patients (or a random sample of the overall population) with a certain disease resident in a given geographic area within a given time period [22–24]. These registries collect a standardized set of variables for every case of the disease in question occurring within a well-defined population. Developed in the first half of the 20th century to provide an understanding of the scale and profile of cancer within communities, and to elucidate causes of variations between and within populations over time, population-based cancer registries in higher-income settings have evolved and frequently measure and assess patterns and quality of care, as well as longitudinal patient outcomes [4,17,25–27].

Clinical registries dedicated to specific cancers exist in the USA, Asia, Australia, and several European countries. Clinical registries collect additional detailed information on diagnostic procedures, pathology examinations, treatment, and follow-up. Importantly, the coding system and
activities for the clinical registry can sometimes be housed and fully integrated in the national population-based incidence registry [28]. Although they usually include a smaller number of patients than population-based cohorts, disease-specific registries may contain more focused and reliable data in some instances. Population-based studies often rely on data sources that are not primarily intended for research, including administrative discharge data and/or billing claims [4]. Conversely, well-designed disease-specific registries define a priori variables needed to assess clinically meaningful outcomes.

Although there are certain differences, these two types of registry can overlap considerably in character and function. Certain disease-specific registries contain data on all the patients diagnosed with PCa in a given geographic area and thus are tantamount to population-based cohorts [15].

3.3. PCa registries

Table 1 lists the characteristics of several currently available population-based cohorts and PCa registries. Among the most prominent population-based cohorts representative of patients from the USA and Europe are the Surveillance, Epidemiology, and End Results (SEER) registry, the linked SEER-Medicare database, and clinical registries in Sweden, Norway, and Japan [25,26,28–32].

Established as part of the National Cancer Act in 1971, the SEER registry provides national cancer statistics that support efforts to reduce the burden of cancer among the US population [4,33]. At present, the SEER program includes data for approximately 28% of the US population. Completeness for case ascertainment is 98%, and approximately 1 000 000 patients with PCa diagnosed between 1973 and 2013 are included in the SEER database. This registry routinely collects data on patient demographics, tumor stage and grade, first course of treatment, vital status, and cause of death (obtained from death certificate review) [4,26,33]. A signed research data agreement form is required to access the SEER data. Several quality control activities are routinely implemented to improve data quality. Potential limitations of the SEER database for PCa include lack of details on patient comorbidities, tumor characteristics, subsequent cancer treatments, and the occurrence of biochemical and clinical progression (Table 2) [4,26].

The linked SEER-Medicare database overcomes some limitations encountered for SEER data alone. This registry derives from the linkage of population-based tumor registry data from the SEER program to healthcare claims from the Medicare program, a US national social insurance program that provides health insurance for Americans aged ≥65 yr. The first linkage was completed in 1991, with subsequent updates approximately every 2 yr. With each linkage, more than 90% of people aged ≥65 yr diagnosed with an incident cancer in one of the SEER catchment areas are matched to claims in the Medicare enrollment file. By virtue of this linkage, the SEER-Medicare database includes claims for hospital, physician, outpatient, home health, and hospice services in addition to the cancer-specific data provided by the SEER database. More than 200 000 patients diagnosed with PCa are currently enrolled in the SEER-Medicare database. Investigators must obtain approval from the National Cancer Institute (NCI) for specific research questions to access the data. Data on comorbidities, as well as details on the type of primary treatment, use of additional cancer therapies, and costs of care, are accessible via analyses of Medicare claims [4,25]. However, these claims are intended for reimbursement and were not collected primarily for clinical research. As a consequence, some clinically relevant details are often lacking. For example, actual prostate-specific antigen (PSA) values after treatment (rather than just receipt of a PSA test) and data on patient-reported quality of life after treatment are not available from standard Medicare claims. In addition, the validity of administrative claims can be substantially limited by inaccurate coding, as well as variability in coding practices among physicians, practices, and hospitals [4,34]. For example, owing to problems with the quality of PSA values at diagnosis, the NCI recently removed these data from the updated SEER-Medicare database. The growing proportion of men enrolled in Medicare-managed care programs—who differ in nonrandom ways from those in Medicare fee-for-service—are also excluded. Finally, results obtained using SEER-Medicare data might not be generalizable to patients younger than 65 yr.

The NCI Prostate Cancer Outcomes Study (PCOS) was initiated in 1994 to investigate cross-sectional patterns of care for PCa and how its treatments impact short- and long-term quality-of-life outcomes [3,4]. Validated questionnaires were sent to approximately 3500 patients diagnosed with PCa in 1994 or 1995 selected from the geographic regions of six SEER registries. The PCOS questionnaire allowed detailed assessments of health-related quality-of-life outcomes (eg, urinary, bowel, and sexual function and bother) that are not available in SEER or SEER-Medicare. Participation was voluntary, and approximately 40% of patients who were invited to join the study elected not to, which may introduce a selection bias. In addition, changes in the diagnosis and treatment of PCa since the inception of PCOS nearly 20 yr ago, as well as the introduction of novel technologies and changes in the Gleason grading system, might limit the applicability of PCOS findings to patients diagnosed today [3,35–37].

More recently, the Comparative Effectiveness Analysis of Surgery and Radiation (CAESAR) study established a prospective population-based cohort of patients with newly diagnosed, clinically localized PCa [15]. The CAESAR cohort includes patients diagnosed in five SEER registries, as well as those included in the CaPSURE program during 2011–2012. Participation in the registry is voluntary; both the treating physician and the patient might refuse to participate. Given its recent introduction, this prospective registry is still limited by relatively short follow-up.

National population-based registries of patients with PCa are also available in Europe. The registries in Norway and Sweden, for example, provide detailed information over and above the basic data items on incidence collected from clinicians and pathologists in the national population-based cancer registry.
<table>
<thead>
<tr>
<th>Name</th>
<th>Registry aim</th>
<th>Patient characteristics</th>
<th>Data collection</th>
<th>Sites included, geographic area, and time period</th>
<th>Patients (n)</th>
<th>Outcomes evaluated</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Population-based registries</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| SEER: Surveillance, Epidemiology, and End Results | To monitor cancer trends and provide data on cancer incidence, extent of disease at diagnosis, therapy, and survival | Population-based registry including patients with incident PCa | Prospective data collection | ~28% of the US population | 1,000,000 | - PCa incidence  
- Initial treatment  
- Disease characteristics at presentation  
- Pathologic characteristics  
- Cancer-specific survival and cause of death |
| SEER-Medicare linked database | To provide detailed data on cancer treatments, health services utilization, and outcomes | Population-based registry including PCa patients >65 yr who are also fee-for-service Medicare beneficiaries | Prospective data collection | Linkage of two population-based data sources: the SEER cancer registry and Medicare claims files | 234,000 | - ICD-9 diagnosis and procedure codes (Medicare)  
- CPT and HCPCS codes (Medicare)  
- Baseline comorbidities based on administrative codes (Medicare)  
- Cancer characteristics (SEER)  
- PCa incidence and mortality (SEER) |
| PCOS: Prostate Cancer Outcomes Study | To investigate how PCa and its treatments affect health-related QoL and to explore patterns of care | Patients diagnosed with PCa between 1994 and 1995 from six SEER regions | Prospective data collection | 6 SEER registries: Connecticut, Utah, New Mexico, and the metropolitan areas of Atlanta, Los Angeles, and Seattle | 3533 | - Sexual, urinary, and bowel function and bother  
- Health-related QoL  
- Disease characteristics  
- Detailed comorbidity information  
- Cancer-specific survival and cause of death |
| CAESAR: Comparative Effectiveness Analyses of Surgery and Radiation | To compare the effectiveness and harms of different treatments for men with localized PCa and assess the relationship of quality of care and outcomes | Patients with newly diagnosed localized PCa | Prospective observational cohort study; patient-reported information collected at baseline and 6 and 12 mo; ongoing data collection at 3 and 5 yr after treatment | Patients from 5 SEER registries (California, Georgia, Louisiana, New Jersey, Utah) and the CaPSURE registry | 3600 | - Treatment  
- Complications  
- Detailed comorbidity information  
- Various psychosocial and personality characteristics of the participants  
- QoL  
- Disease recurrence  
- Quality of care information |
| PCaSe: Prostate Cancer data Base Sweden | To provide a platform for PCa research in Sweden | Population-based registry including PCa patients | Prospective data collection | National Prostate Cancer Register of Sweden linked to national registers; 1996–2009 | 120,000 | - Inpatient and outpatient care  
- Patterns of use of therapies  
- Use of prescribed drugs  
- Socioeconomic and family factors  
- Detailed follow-up information |
| Prostate Cancer Clinical Registry (part of the Cancer Registry of Norway) | To provide data for monitoring inpatient and outpatient outcome and survival | Population-based registry including PCa patients | Prospective data collection | National Prostate Cancer Register of Norway linked to national registry from 2009 | 30,000 | - Prognostic factors  
- Treatment outcomes  
- Evaluation of the quality of PCa care |
| J-CaP: Japanese Study Group of Prostate Cancer | To assess the outcomes of hormone therapy among PCa patients | PCa patients treated with androgen deprivation therapy | Prospective data collection | Japanese patients receiving androgen deprivation therapy from 2001 | 26,000 | - Prognostic factors  
- Treatment outcomes |
| **Prospective patient registries** | | | | | | |
| CaPSURE: UCSF Cancer of the Prostate Strategic Urologic Research Endeavor | To expand knowledge of risk prediction, practice trends, outcomes, costs, and QoL for PCa | Patients with all stages of biopsy-proven PCa | Longitudinal data collection starting at baseline and continuing every 6 mo | 47 community urologic practices, academic medical centers, and VA sites across the USA since 1995 | 15,000 | - Preoperative characteristics  
- Diagnostic workup  
- Initial and subsequent treatments  
- Pathologic and oncologic outcomes  
- QoL after treatment  
- General health  
- Resource utilization |
<table>
<thead>
<tr>
<th>Name</th>
<th>Registry aim</th>
<th>Patient characteristics</th>
<th>Data collection</th>
<th>Sites included, geographic area, and time period</th>
<th>Patients (n)</th>
<th>Outcomes evaluated</th>
</tr>
</thead>
</table>
| **MUSIC: Michigan Urological Survey Improvement Collaborative** | To evaluate and improve the quality and cost efficiency of PCa care for men in Michigan | All patients undergoing a prostate biopsy in participating practices, and all patients seen for newly diagnosed PCa | Prospective data collection | 42 urology practices from throughout the Michigan including >80% of urologists in the state; data collection began in March 2012 | 15,000 | - Imaging use  
- Use and complications of prostate biopsy  
- Patient-reported outcomes after surgery  
- Variation in patterns of care for men with early-stage cancers |
| **SA-PCCOC: South Australian Prostate Cancer Clinical Outcome Collaborative** | To evaluate the standard of care and outcomes and improve quality of care for men with PCa in South Australia | Patients with a new histological diagnosis of PCa or treatment for this disease | Prospective longitudinal data collection | Patients from the major metropolitan and teaching hospitals and collaborating private practitioners and institutions in Adelaide, South Australia; 1998-2015 | 9,500 | - Pretreatment clinical, pathologic, and patient-reported Qol data  
- Pathologic outcomes after surgery  
- Treatment details and complications  
- Follow-up data including PSA, clinical evidence of recurrence, symptoms, and Qol data  
- Mortality: all-cause and PCa-specific |
| **Victorian Prostate Cancer Registry (Australia)** | To monitor the patterns of care and outcomes of men diagnosed with prostate cancer in Victoria | Patients with newly diagnosed PCa | Prospective data collection with EPIC administered at 12 and 24 mo | Patients from public and private hospitals, capturing >90% of all new PCa diagnoses | >10,000 | - Preoperative characteristics  
- Management details  
- Positive surgical margins  
- Biochemical recurrence  
- Additional cancer therapies  
- Death  
- Qol assessed at 12 and 24 mo using EPIC |
| **K-CaP: Korean Prostate Cancer Database** | To analyze clinical and pathologic PCa outcomes to improve patient care | Patients with newly diagnosed PCa | Prospective data collection | Patients from 5 Korean institution diagnosed since 2011 | 858 | - Disease characteristics at presentation  
- Oncologic outcomes  
- Functional data assessed using validated questionnaires |
| **AQUA: AUA Quality Registry** | To improve the quality of care of patients with urologic diseases | Patients with newly diagnosed PCa | Patients will be followed prospectively from the time of diagnosis | 100 sites in the USA by the end of 2016 | NA | - Quality of documentation  
- Quality of care (process measures)  
- Patient-reported Qol outcomes |

ICD = International Classification of Diseases; CPT = current procedural terminology; HCPCS = Healthcare Common Procedure Coding System; QoL = quality of life; UCSF = University of California at San Francisco; VA = Veterans Administration; PSA = prostate-specific antigen; EPIC = Expanded Prostate Cancer Index Composite; NA = not available.
<table>
<thead>
<tr>
<th>Registry</th>
<th>Strengths</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Population-based registries</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SEER: Surveillance, Epidemiology, and End Results</td>
<td>• Large population-based cohort representative of the entire US population</td>
<td>• No data on comorbidity</td>
</tr>
<tr>
<td></td>
<td>• Detailed data on disease characteristics</td>
<td>• Limited details on the type of treatment</td>
</tr>
<tr>
<td></td>
<td>• Relatively long follow-up</td>
<td>• No information on treatment complications</td>
</tr>
<tr>
<td></td>
<td>• Data on vital status and cause-of-death</td>
<td>• No data on adjuvant or salvage therapies</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Lack of information on biochemical and/or clinical recurrences</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• No details on radiotherapy doses or surgical technique</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• No patient-reported outcomes</td>
</tr>
<tr>
<td>SEER-Medicare linked database</td>
<td>• Large population-based cohort representative of the US population aged &gt;65 yr</td>
<td>• Inclusion of only Medicare beneficiaries aged &gt;65 yr</td>
</tr>
<tr>
<td></td>
<td>• Detailed data on disease characteristics</td>
<td>• Use of administrative codes and billing information other than clinical data</td>
</tr>
<tr>
<td></td>
<td>• Relatively long follow-up</td>
<td>• Lack of specific information on biochemical and/or clinical recurrence</td>
</tr>
<tr>
<td></td>
<td>• Details on comorbidity and treatment complications</td>
<td>• No patient-reported outcomes</td>
</tr>
<tr>
<td></td>
<td>• Data on treatment type and receipt of adjuvant and/or salvage therapies</td>
<td>• Problems with prostate-specific antigen data points resulting in National Cancer Institute advising investigators not to use this variable</td>
</tr>
<tr>
<td>PCaSOS: Prostate Cancer Outcomes Study</td>
<td>• Use of validated questionnaires for patient-reported outcomes</td>
<td>• Relatively small sample size</td>
</tr>
<tr>
<td></td>
<td>• Patients sampled according to a prespecified design to ensure a representative sample with ethnic and racial diversity</td>
<td>• Inclusion of patients diagnosed within a limited historic time period (1994–1995)</td>
</tr>
<tr>
<td>CAESAR: Comparative Effectiveness Analyses of Surgery and Radiation</td>
<td>• Extensive health-related quality of life assessment administered at baseline and 6 and 12 mo after enrollment</td>
<td>• Surgical and radiation techniques have evolved in the past 15 yr, so findings may not be generalizable to current patients</td>
</tr>
<tr>
<td></td>
<td>• Inclusion of nontraditional patient characteristics that appear to improve risk adjustment</td>
<td>• There may be some selection bias among the study participants, as only half of the patients approached to participate agreed to join the study</td>
</tr>
<tr>
<td></td>
<td>• Assessment of quality-of-care measures</td>
<td></td>
</tr>
<tr>
<td>PCAxBSE: Prostate Cancer data Base Sweden/Prostate Cancer Clinical Registry (part of the Cancer Registry of Norway)</td>
<td>• Nationwide cohort including virtually all PCA patients in Sweden</td>
<td>• Relatively short follow-up</td>
</tr>
<tr>
<td></td>
<td>• Detailed follow-up data</td>
<td>• Represents only selected geographic areas of the USA</td>
</tr>
<tr>
<td></td>
<td>• Information on comorbidity</td>
<td>• There may be some selection bias among the study participants, as only half of the patients approached to participate agreed to join the study</td>
</tr>
<tr>
<td>J-CaP: Japanese Study Group of Prostate Cancer</td>
<td>• Nationwide longitudinal prospective cohort study</td>
<td>• Only PCA patients receiving androgen deprivation therapy included in the registry</td>
</tr>
<tr>
<td></td>
<td>• Detailed data on type of androgen deprivation therapy</td>
<td></td>
</tr>
<tr>
<td><strong>Prospective patient registries</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CaPSURE: UCSF Cancer of the Prostate Strategic Urologic Research Endeavor</td>
<td>• Large cohort of patients with all stages of PCA</td>
<td>• Urology practices over-represented and radiology practices under-represented</td>
</tr>
<tr>
<td></td>
<td>• Collection of both clinician- and patient-reported outcomes</td>
<td>• Only diagnostic and therapeutic studies ordered by participating physicians are recorded</td>
</tr>
<tr>
<td></td>
<td>• Validated instruments to assess patient outcomes collected at baseline and longitudinally</td>
<td></td>
</tr>
<tr>
<td>MUSC: Michigan Urological Survey Improvement Collaborative</td>
<td>• Inclusion of the majority of urologists in Michigan state</td>
<td>• Relatively short follow-up</td>
</tr>
<tr>
<td></td>
<td>• Assessment of validated patient-reported outcomes</td>
<td>• Represents only the practice of urologists in the state of Michigan</td>
</tr>
<tr>
<td></td>
<td>• Evaluation of quality of care measures</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Tri-annual collaborative-wide meetings</td>
<td></td>
</tr>
<tr>
<td>SA-PCCOC: South Australian Prostate Cancer Outcome Collaborative</td>
<td>• Patient characteristics broadly representative of all South Australian men with PCA diagnosis</td>
<td>• Small number of patients with very long-term data available</td>
</tr>
<tr>
<td></td>
<td>• Large cohort of patients inclusive of all treatments and all stages of PCa and detailed patient-reported outcomes</td>
<td>• Imbalance in inclusion between patients treated in public and private hospitals</td>
</tr>
<tr>
<td></td>
<td>• Electronic and third-party data collection independent of treating physician(s)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Detailed data on disease characteristics and quality of life at baseline and forward from diagnosis</td>
<td></td>
</tr>
<tr>
<td>Victorian Prostate Cancer Registry (Australia)</td>
<td>• Data capture almost at population level (&gt;90%) for a large region</td>
<td>• Relatively short follow-up</td>
</tr>
<tr>
<td></td>
<td>• Detailed data on initial presentation and management</td>
<td>• Small numbers of participating hospitals initially</td>
</tr>
<tr>
<td></td>
<td>• Feedback to clinicians of their outcomes using quality indicators (deidentified funnel plots)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Third-party administration of quality of life questionnaire</td>
<td></td>
</tr>
</tbody>
</table>
The Prostate Cancer Data Base of Sweden (PcBaSe) registry includes more than 110 000 patients diagnosed with PCa between 1999 and 2009. The PcBaSe registry derives from a linkage between the National Prostate Cancer Registry of Sweden and several other nationwide registries [28]. The main strengths of this database are its automatic inclusion of virtually all patients diagnosed with PCa in Sweden and the availability of detailed follow-up information.

At the Cancer Registry of Norway, the clinical PCa registry was established in 2009 and includes registration of treatment and follow-up data. All medical doctors in Norway are instructed by law to notify new cases to the registry. The aims are to provide data for monitoring patient outcomes and survival, and to serve as an empirical base for scientific studies concerning prognostic factors, treatment outcomes, and evaluation of the quality of cancer care. The registry has a reference group comprising a panel of multidisciplinary experts from clinical and research domains within the country who advise on the operation of the clinical registry and its strategic direction.

Finally, the Japanese Study Group of Prostate Cancer (J-CaP) established a nationwide longitudinal prospective cohort study in 2001 to evaluate the outcomes for patients with PCa undergoing androgen deprivation therapy. The registry includes data from eligible institutions participating on a voluntary basis. Overall, more than 26 000 patients were enrolled in the study [32]. Nonetheless, it should be noted that this registry focuses on men treated with hormonal therapies.

In addition to these population-based registries, several well-designed prospective patient registries have been developed [3]. Cancer of the Prostate Strategic Urologic Research Endeavor (CaPSURE) represents one of the first disease-specific longitudinal registries for PCa patients. Since its inception in 1995, CaPSURE has grown to include data for more than 15 000 patients from 47 urology practices, mostly community-based but also including four academic medical centers and Veterans Administration hospitals [38]. The registry is managed by the Urology Department at the University of California, San Francisco. The main aims of this registry are to examine trends in diagnosis and management, clinical outcomes, health-related quality of life, and resource utilization among PCa patients. Clinical information on patients who have consented to be included in the study is collected by the treating physicians in participating centers at baseline and every 6 mo during follow-up. In addition, patient-reported questionnaires are administered at enrollment and during follow-up visits. CaPSURE also recently planned to start collecting biospecimens from both prostate biopsies and radical prostatectomies [3], thereby allowing future analyses that combine these biomarkers with clinical data for the development of novel prognostic tools. Although the availability of detailed clinical data (~1000 variables in total) for a large number of patients with all stages of PCa makes this database unique, urology practices are over-represented in comparison to radiation practices, and this could limit the generalizability of some analyses from this registry [4].

The Michigan Urological Surgery Improvement Collaborative (MUSIC), a physician-led collaborative quality improvement initiative supported by Blue Cross Blue Shield of Michigan, includes data for patients with PCa diagnosed and treated by more than 200 physicians at 42 urology practices across the state of Michigan [39]. The collaborative is designed to improve the quality and cost efficiency of PCa management. Practices included in the collaborative voluntarily submit demographic and clinical data on all patients with newly diagnosed PCa. Many of the practices also collect data on patient-reported functional outcomes after radical prostatectomy. The priority areas for MUSIC include appropriate imaging, increasing the safety of transrectal ultrasound-guided prostate biopsies, enhancing outcomes after surgery, and optimizing treatment decisions for men with newly diagnosed PCa [39].

The South Australian Prostate Cancer Clinical Outcome Collaborative (SA-PCCOC) is a database established in 1998 that includes all patients with a histologic diagnosis of PCa in the major public and teaching hospitals in the state of South Australia and a number of collaborating private practitioners and private institutions [40]. SA-PCCOC participation is complete in public hospitals and voluntary in the private sector. Approximately 9500 patients are currently enrolled. The registry prospectively captures 90% of all new PCa diagnoses in South Australia. However, as it collects >95% of all PCa diagnoses in the public health system but only ~80% of those in the private sector, reported outcomes may be biased by treatment differences between these two systems. The objective of the database is to evaluate the standard of care for men with PCa in South Australia by monitoring the outcomes of care over time. SA-PCCOC is a multidisciplinary, longitudinal, prospective
registry. Information is collected independent of the treating physician(s) via electronic methods and trained third-party data collectors. SA-PCCOC has extensive high-quality PSA data for all patients, including values before entry extending back several years, as well as access to tissue biobanks.

Also in Australia, the Victorian Prostate Cancer Registry (PCR), established in 2008, now captures >90% of all new PCa diagnoses in Victoria, the second most-populous state in Australia. Only clinicians working within contributing hospitals are eligible to enroll in the PCR on a voluntary basis. The PCR captures a detailed data set at inclusion to allow risk-stratification using the NCCN and CAPRA systems, as well as initial disease management information [41]. Participating clinicians receive regular reports summarizing their contribution and reporting their performance for a range of quality indicators (eg, positive surgical margins, urinary bother, sexual bother) in comparison to deidentified colleagues. The PCR also includes the Expanded Prostate Cancer Index Composite (EPIC) questionnaire at 12 and 24 mo. As of late 2014, it had enrolled >10 000 patients and has led to significant publications reporting patterns of care [42], positive surgical margins [43], and the use of active surveillance [44]. As a result, this Melbourne-based registry has been awarded a grant by the Movember Foundation to extend its reach by creating a national PCa registry for Australia.

The Korean Prostate Cancer Database (K-CaP) has recently been started and includes data on 858 patients with PCa treated with radical prostatectomy at three institutions. Participation in the registry is on a voluntary basis, and eligible institutions include all urologic centers in Korea. Although this registry includes a small number of patients with short follow-up, it represents the first comprehensive database for PCa aimed at improving the quality of PCa in Korea [45].

Finally, the American Urological Association (AUA) recently launched a new registry for patients with PCa diagnosed in the USA. The AUA Quality (AQUA) registry is designed to measure and report quality of care and outcomes for patients with PCa and other urologic conditions. This registry will prospectively collect data from approximately 100 sites by the end of 2016. The AQUA registry will allow for automated collection of clinical data directly from electronic medical records. Participation in the registry is voluntary; however, the main long-term goal is to include every US urologic practice. The introduction of a national prospective patient registry would result in the availability of a large amount of data, with findings potentially generalizable to the US population [3].

3.4. The value of PCa registries

Table 3 presents examples of clinical investigations that can be performed using PCa registries. Historically, population-based studies were performed to assess incidence and survival of a certain disease in a particular geographic area [23,46,47]. In the context of PCa, each year the SEER registry provides data on the numbers of new cases diagnosed and deaths that occurred in the current year in the USA [46,47]. From population-based registries, data on the baseline characteristics of patients included might be used to analyze temporal trends in PCa presentation and risk [48–51]. In the five Nordic countries, an analysis of incidence and mortality data revealed a rapid increase in PCa incidence during the early 1990s coinciding with the introduction of PSA testing, while mortality rates stabilized or declined in countries where PSA testing and curative treatment have been commonly practiced since the late 1980s [52]. In addition, tumor registries can provide incidence and mortality data on patients with rare histologic subtypes that are not included in RCTs or institutional cohorts [53]. Cancer registries are also used to examine differences in incidence and mortality rates according to patient characteristics such as age, comorbidity status, race, socioeconomic status, and disease severity [54–59]. Potential risk factors and prognostic factors can be identified from these data as well [13]. In addition, they might be useful for assessment of trends in the adoption of screening, diagnostic, and treatment procedures, as well as novel technologies [54,56,58,60–67]. For example, Vickers et al [68] demonstrated that restriction of PSA testing to only young men or selected men aged >70 yr might reduce the risk of overdiagnosis. Conversely, Hu et al [69] showed that in men aged >70 yr included in the SEER-Medicare database, the frequency of PSA screening in the 5 yr before diagnosis was associated with a lower risk of harboring metastatic disease and with higher overall- and cancer-specific survival.

A recent study performed using the MUSIC registry demonstrated that adoption of active surveillance in low-risk patients varies widely among urology practices in the same state [66]. By linking data from the Cancer Registry of Norway to the incidence of definitive radiotherapy or radical prostatectomies, Kvale et al [70] reported that the earliest declines in PCa mortality were seen in Norwegian regions where curative treatment was most frequently used [70]. Disparities and variations in the use of imaging and treatment modalities can also be evaluated using disease-specific registries. In particular, these databases allow examination of the relationship between baseline characteristics and the likelihood of receiving proper staging or adequate treatment [54,56,58,60–67]. A recent CaPSURE study demonstrated substantial differences in primary treatment between African-American and white men with similar risk profiles [54].

Data from registries can provide important insights into clinical guideline adherence among physicians practicing in a certain geographic area or included in a prospective registry [71–75]. Chen et al [72] demonstrated that substantial discordance exists in the SEER-Medicare regarding adoption of the National Comprehensive Cancer Network guidelines for patients with high-risk PCa according to age at diagnosis. Similarly, registries can provide information on the quality of clinical care. In this context, studies based on several of these data sources have demonstrated that better outcomes might be achieved in high-volume hospitals and/or with more experienced
surgeons [21,76]. The evaluation of large cohorts with follow-up data on cancer control has an important role in assessment of the safety and efficacy of PCA treatments [77–94]. Data on health-related quality of life after treatment can also be analyzed. Sexual dysfunction and urinary incontinence and/or irritative symptoms are common long-term sequelae after PCA treatment [35,55]. Since the incidence of erectile dysfunction and urinary incontinence may vary substantially according to definition, the use of patient-reported outcomes and validated questionnaires is preferred [96–100]. In this context, many prospective disease-specific registries adopt validated instruments to measure health-related outcomes at baseline and during follow-up [15,35,38,39]. This in turn allows more unbiased comparisons of patient outcomes after different treatments [35,38].

Another important endpoint for patients with PCa is diagnosis- and treatment-associated expenditures [101].

### Table 3 – Clinically relevant studies using prostate cancer (PCa) registries

<table>
<thead>
<tr>
<th>Registry</th>
<th>Clinically relevant studies</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Population-based registries</strong></td>
<td></td>
</tr>
</tbody>
</table>
| SEER: Surveillance, Epidemiology, and End Results | - Incidence and mortality of PCa in the USA [27,46]  
- Treatment patterns and racial differences among patients diagnosed with PCa [63]  
- Efficacy of curative-intent treatments among patients with metastatic PCa [87]  
- Adoption and extent of pelvic lymph node dissection among patients with PCa [74]  
- PCa-specific and other-cause mortality according to primary treatment [77] |
| SEER-Medicare linked database | - Comparative effectiveness of open versus robot-assisted RP [78]  
- Impact of primary ADT on survival in men with clinically localized PCa [119]  
- ADT and the risk of cardiovascular diseases, diabetes, fractures, and acute kidney injury in PCa patients [120,121]  
- Racial disparities and geographic variations in primary treatments among PCa patients [64]  
- Increased use of advanced treatment technologies among patients with low-risk PCa and high risk of non-cancer-related mortality [122]  
- Cost implications of the introduction and widespread adoption of advanced treatment technologies for PCa [103] |
| PCOS: Prostate Cancer Outcomes Study | - Association between age, race, and comorbidity and the probability of receiving a conservative treatment [67]  
- Comparative effectiveness of radical prostatectomy and radiation therapy [37]  
| CAESAR: Comparative Effectiveness Analyses of Surgery and Radiation | - Understanding baseline sexual, urinary, and bowel function among contemporary men with newly diagnosed PCa [123]  
- Differences in pretreatment urinary and sexual function in PCa patients enrolled nearly 20 yr apart [124] |
| PCaSe: Prostate Cancer data Base Sweden | - Patterns of ADT administration among patients diagnosed with PCa [65]  
- Psychological distress following a prostate cancer diagnosis [125]  
- Endocrine treatment and the risk of nonfatal and fatal cardiovascular diseases among PCa patients [126] |
| Prostate Cancer Clinical Registry (part of the Cancer Registry of Norway) | - National diagnostic and therapeutic interventions available to Norwegian patients with PCa in the context of EAU guidelines [127] |
| J-ScP: Japanese Study Group of Prostate Cancer | - Treatment and 5-yr survival in patients with nonmetastatic PCa [128] |
| **Prospective patient registries** | |
| CaPSURE: UCSF Cancer of the Prostate Strategic Urologic Research Endeavor | - Development and validation of the CAPRA score to predict pathologic and biochemical outcomes after RP [129]  
- Time trends and variation in primary treatment of PCa in the USA [130]  
- Long-term health-related quality-of-life outcomes after PCa treatment [95]  
- Comparative effectiveness of RP, radiotherapy, and ADT in patients with clinically localized PCa [89]  
- Impact of ADT on physical, mental, and emotional wellbeing in PCa patients [131]  
- Association between race and the likelihood of upgrading, upstaging, and positive surgical margins at RP [132]  
- Cost burden of PCa according to treatment type [106] |
| MUSIC: Michigan Urological Survey Improvement Collaborative | - Guideline adherence for radiographic staging of men with newly diagnosed PCa [133]  
- Implementation of novel clinical pathways to address fluorouracil-resistant and infection-related hospitalizations among men undergoing prostate biopsy [134]  
- Implementation of performance feedback and collaborative-wide educational activities to improve documentation of clinical TNM stage in PCa patients [135]  
- Variation in the use of active surveillance for men with low-risk PCa [66] |
| SA-PCOCC: South Australian Prostate Cancer Clinical Outcome Collaborative | - Association between postoperative PSA kinetics and cancer-specific mortality after external beam radiotherapy [136]  
- Accuracy of previously validated risk assessment tools in Australian patients with PCa [137] |
| Victorian Prostate Cancer Registry (Australia) | - Contemporary patterns of care for men newly diagnosed with PCa [42]  
- Positive surgical margins: rate, contributing factors and rate of additional therapy [43]  
- Active surveillance patterns of care in a contemporary PCa registry [44] |

**RP** = radical prostatectomy; **ADT** = androgen deprivation therapy; **EAU** = European Association of Urology; **CAPRA** = Cancer of the Prostate Risk Assessment; **PSA** = prostate-specific antigen.
Although PCa-related costs vary widely across different health care systems, large administrative databases can in many cases provide an accurate estimate of the economic burden of PCa, including the cost implications of screening programs and the introduction of novel technologies [102–106].

Finally, data from registries can be used as the foundation for quality improvement activities in the management of PCa patients [107]. Population-based studies are valuable in determining the average level of care in a certain geographic region, allowing identification of areas where interventions are needed to improve the quality of care [70,108–110]. In addition, prospective disease-specific registries collect important data on practice patterns, processes of care, and validated outcomes [3,15,39]. If these data are shared directly with the practices and the results are compared between physicians and hospitals, providers are likely to make substantial efforts to improve their results [39,107]. Knowledge of such comparative performance feedback motivates providers to adopt quality improvement measures that eventually improve patient outcomes.

3.5. Pitfalls of current PCa registries

PCa registries are not without limitations that should be considered when performing and reporting analyses based on these data. The main limitation reflects the observational, nonrandomized nature of the data. Although several statistical modeling techniques can be applied to increase the validity of results obtained, unmeasured confounding variables and selection biases remain legitimate threats to the validity of such studies [25,111]. As mentioned previously, a second limitation is that for some population-based registries, the available data were not originally collected for research purposes [7]. Moreover, diagnosis and procedure codes can be incomplete or even inaccurate in some cases. Procedures to monitor and improve data quality are a necessary component of these registries. Third, population-based registries usually include patients diagnosed over a long time period. Therefore, changes in staging and grading classification systems, as well as modifications for coding systems, can result in misclassification of some variables [112]. These issues should be considered when planning investigations using these registries.

For prospective disease-specific registries, an important limitation can be the relatively small number of patients included. In addition, because participation in these programs requires substantial resources, the sites are often not randomly chosen and results from studies performed using prospective registries may not be applicable to the entire population [4]. Ultimately, enrolment of patients on a voluntary basis, which is typical of some population-based and prospective registries, might introduce selection biases and in turn limit the generalizability of findings obtained in this setting to the entire population. This is particularly true for registries with low response rates, for which only individuals with better outcomes might be included in these studies [15]. Each of these limitations should be carefully considered when researchers design, implement, and report studies based on these databases.

3.6. Statistical methods to increase the generalizability of results from PCa registries

As mentioned previously, selection bias and unmeasured confounding variables can limit the validity of findings obtained from population-based and disease-specific registries [111]. Over the last decades, several methodological advances have been introduced that can address these limitations in some circumstances.

Standard multivariable analyses can be used to estimate the independent association between an exposure and an outcome after adjusting for measured confounding variables. As an additional step, propensity score matching might be applied. This statistical technique attempts to estimate the effect of a treatment by accounting for factors that predict the receipt of the treatment itself, allowing for identification of a control group that is better matched with treated subjects for available covariates. This in turn potentially reduces bias in estimates of the effect of a treatment [113]. Although propensity score matching better balances measured confounders, previous studies hypothesized that forced balance of measured confounders might exacerbate the imbalance in unmeasured covariates. If unmeasured covariates are confounders, propensity score matching might aggravate the selection bias [114–116]. Therefore, some caution is needed when using this statistical approach.

Neither multivariable analyses nor propensity score approaches can address residual confounding and/or selection bias due to unmeasured variables. In some cases, instrumental variable analysis can be used as a technique for balancing unmeasured confounding variables, thereby yielding unbiased estimates of treatment effects. Conceptually, instrumental variable analysis is used to achieve pseudorandomization between two treatment groups, thereby balancing both measured and unmeasured covariates [116]. This technique uses a variable, called an instrument, that affects treatment choice but is not related to the outcome except through the choice of treatment [117]. This allows determination of the level of exogenous variation, for example, how the treatment variable affects outcome. Indeed, variations in treatment resulting from variations in the value of the instrument are considered analogous to variations in treatment resulting from randomization. In this context, previous studies demonstrated that the adoption of instrumental variable analyses provided similar results to observations in clinical trials [115]. Nonetheless, the value of instrumental variable analyses strongly depends on the quality of the instrument, and such analyses should be undertaken only in collaboration with an experienced biostatistician and/or econometrician.

3.7. Future directions

Population-based and disease-specific PCa registries will continue to accrue patients, so progressively larger cohorts will be available in the future. Although these databases
provide unique information regarding disease characteristics, patterns of care, and patient outcomes, there are several steps that could further enhance the value of clinical research based on these data sources (Table 4).

First, collection of high-quality, clinically relevant data is mandatory and should ideally be prospective in nature. In addition, clinically meaningful covariates should be identified a priori according to the main aims of the registry [15]. Quality control and improvement initiatives should be conducted to reduce potential biases related to data acquisition and/or coding procedures. These might include data audits and validation studies based on external data sets [33,138]. Finally, continuous training of registry personnel and data abstractors to ensure standardization of data collection procedures is desirable, although admittedly resource-intensive.

Second, PCa registries should ideally produce reports and data that provide clinicians with comparative performance feedback [39,107]. The availability of such data could motivate many physicians and hospitals to improve their daily clinical practice and foster innovations that yield new knowledge and best practices in the field [139]. However, these data could have significant unintended consequences if they are used for marketing or in a punitive fashion by payers or other stakeholders [39].

Third, in the emerging era of precision medicine, disease-specific registries should ideally incorporate tissue- or serum-based biomarker data. Collection of urine and serum samples, as well as biopsy and prostatectomy specimens, would allow identification and validation of novel models to predict long-term outcomes for patients with PCa. Correlation between biomarker data and PCa outcomes among large patient samples included in these registries might improve our understanding of the natural history of the disease.

Finally, although RCTs might be difficult to initiate and complete because of their high costs, substantial expense associated with data collection and quality assurance processes might limit the creation and widespread availability of disease-specific registries [15,38,39,107]. Some registries are currently funded by industry and federal grants. Alternatively, payers might be involved in the financial support of registries. The business case for this funding is based on the notion that improvements in the quality of care can yield substantial savings in health care expenditure. Supporting this point, prior work has demonstrated that a reduction of only 2% in treatment-related complications yields net savings for payers that support such quality improvement initiatives [118]. As a consequence, there appears to be a strong rationale for greater payer support of initiatives aimed at improving the quality of care and long-term outcomes of patients affected by PCa.

### Table 4 – Potential future directions for prostate cancer registries

<table>
<thead>
<tr>
<th>Ideal characteristics of prostate cancer registries</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Inclusion of patient cohorts representative of the entire population</td>
</tr>
<tr>
<td>• Inclusion of disease-specific covariates and endpoints</td>
</tr>
<tr>
<td>• Use of validated instruments to collect patient-reported preoperative and follow-up data</td>
</tr>
<tr>
<td>• Collection of biopsy and prostatectomy biospecimens</td>
</tr>
<tr>
<td>• Collection of details on the type and quality of treatment administered</td>
</tr>
<tr>
<td>• Evaluation of clinically relevant endpoints</td>
</tr>
<tr>
<td>• Extensive patient surveys and medical chart review to control for potential confounders</td>
</tr>
<tr>
<td>• Sufficient follow-up for clinically-relevant endpoints</td>
</tr>
<tr>
<td>• Collection of data on costs of treatment</td>
</tr>
<tr>
<td>• Implementation of measures for quality control of the collected data</td>
</tr>
<tr>
<td>• Data sharing with participants as a means for quality improvement</td>
</tr>
<tr>
<td>• Independent funding</td>
</tr>
</tbody>
</table>

4. Conclusions

Several population-based and prospective disease-specific registries are available for PCa. Although RCTs still provide the highest level of evidence, analyses of data from these registries play an important role in advancing PCa care. One of the main advantages of disease-specific registries is the possibility of assessing PCa incidence and mortality, disease characteristics, trends in the utilization of health care services, quality of care, long-term outcomes, and costs in real-world practice settings. Nonetheless, some limitations should be carefully considered when planning studies using these databases. Their retrospective nature, inaccurate coding, missing data, possible selection biases related to cost issues, and the voluntary nature of some registries might preclude the generalization of findings obtained in this context. Moving forward, improvements in data quality, collection of tissue samples, and the availability of data for performance feedback and quality improvement will increase the clinical relevance and impact of studies based on data available from these valuable sources.

**Author contributions:** Giorgio Gandaglia had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

**Study concept and design:** Gandaglia, Bray, Cooperberg, Karnes, Leveridge, Moretti, Murphy, Penson, Miller.

**Acquisition of data:** Gandaglia, Bray, Cooperberg, Karnes, Leveridge, Moretti, Murphy, Penson, Miller.

**Analysis and interpretation of data:** Gandaglia, Bray, Cooperberg, Karnes, Leveridge, Moretti, Murphy, Penson, Miller.

**Drafting of the manuscript:** Gandaglia, Bray, Cooperberg, Karnes, Leveridge, Moretti, Murphy, Penson, Miller.

**Critical revision of the manuscript for important intellectual content:** Bray, Cooperberg, Karnes, Leveridge, Moretti, Murphy, Penson, Miller.

**Statistical analysis:** None.

**Obtaining funding:** None.

**Administrative, technical, or material support:** None.

**Supervision:** None.

**Other:** None.

**Financial disclosures:** Giorgio Gandaglia certifies that all conflicts of interest, including specific financial interests and relationships and affiliations relevant to the subject matter or materials discussed in the manuscript (eg, employment/affiliation, grants or funding, consultancies, honoraria, stock ownership or options, expert testimony, royalties, or patents filed, received, or pending), are the following: None.

**Funding/Support and role of the sponsor:** None.


