

# Danish Prostate Cancer Registry – methodology and early results from a novel national database

JT Helgstrand<sup>1</sup>  
 N Klemann<sup>1</sup>  
 MA Røder<sup>1</sup>  
 BG Toft<sup>2</sup>  
 K Brasso<sup>1</sup>  
 B Vainer<sup>2</sup>  
 P Iversen<sup>1</sup>

<sup>1</sup>Copenhagen Prostate Cancer Center, Department of Urology,

<sup>2</sup>Department of Pathology, Rigshospitalet, Copenhagen University Hospital, University of Copenhagen, Copenhagen, Denmark



Correspondence: JT Helgstrand  
 Copenhagen Prostate Cancer Center,  
 Department of Urology, Rigshospitalet,  
 Copenhagen University Hospital,  
 University of Copenhagen, Ole  
 Maaløes vej 24, Section 7521, DK-2200  
 Copenhagen N, Denmark  
 Tel +45 2834 0507  
 Fax +45 3545 2726  
 Email [helgstrand1@me.com](mailto:helgstrand1@me.com)

**Background:** Systematized Nomenclature of Medicine (SNOMED) codes are computer-processable medical terms used to describe histopathological evaluations. SNOMED codes are not readily usable for analysis. We invented an algorithm that converts prostate SNOMED codes into an analyzable format. We present the methodology and early results from a new national Danish prostate database containing clinical data from all males who had evaluation of prostate tissue from 1995 to 2011.

**Materials and methods:** SNOMED codes were retrieved from the Danish Pathology Register. A total of 26,295 combinations of SNOMED codes were identified. A computer algorithm was developed to transcode SNOMED codes into an analyzable format including procedure (eg, biopsy, transurethral resection, etc), diagnosis, and date of diagnosis. For validation, ~55,000 pathological reports were manually reviewed. Prostate-specific antigen, vital status, causes of death, and tumor-node-metastasis classification were integrated from national registries.

**Results:** Of the 161,525 specimens from 113,801 males identified, 83,379 (51.6%) were sets of prostate biopsies, 56,118 (34.7%) were transurethral/transvesical resections of the prostate (TUR-Ps), and the remaining 22,028 (13.6%) specimens were derived from radical prostatectomies, bladder interventions, etc. A total of 48,078 (42.2%) males had histopathologically verified prostate cancer, and of these, 78.8% and 16.8% were diagnosed on prostate biopsies and TUR-Ps, respectively.

**Future perspectives:** A validated algorithm was successfully developed to convert complex prostate SNOMED codes into clinical useful data. A unique database, including males with both normal and cancerous histopathological data, was created to form the most comprehensive national prostate database to date. Potentially, our algorithm can be used for conversion of other SNOMED data and is available upon request.

**Keywords:** database, prostate cancer, epidemiology, registry, SNOMED, histopathology

## Introduction

A steep increase in incidence from 1,408 new cases in 1995 to 4,280 in 2013, as well as a rapid increment in expensive treatment options for both localized and advanced prostate cancer (PCa), makes the malignancy a major challenge for health care systems.<sup>1</sup> As in the rest of the Western world, the incidence of PCa in Denmark has increased rapidly by ~7.2% per year since 1995, and the age-standardized annual incidence has increased from 70 to 163 cases per 100,000 males in the period 1995–2010.<sup>2</sup> Increased awareness and the use of opportunistic prostate-specific antigen (PSA) screening are the primary contributing factors for this development.<sup>3</sup> Still, the PCa-specific mortality has remained unchanged in Denmark at an age-standardized mortality rate of 19 per 100,000.<sup>4</sup>

The impact of these dramatic epidemiological changes on histological grade, stage, and diagnostic and therapeutic activity is not fully appreciated. We initiated a program to create a comprehensive integrated database including all Danish males who underwent histopathological evaluation of prostate tissue in the period 1995–2011 using the national Danish Pathology Register (Patobank) as a platform.<sup>5</sup> Patobank registers diagnoses of all pathology assessments made by Danish pathologists. The diagnoses are given as text and/or Systematized Nomenclature of Medicine (SNOMED) codes, of which ~7,000 different codes were identified. In the individual patient, the SNOMED codes are typically combined to achieve optimal accuracy. However, this results in a huge number of combinations, essentially making raw SNOMED codes impossible to analyze. No previous computational attempts to convert SNOMED codes to analyzable data have been made. With extensive transcoding of Patobank data, we developed a unique algorithm that transformed complex codes to a relevant conclusion including tissue type, histopathological diagnosis, and histopathological grading. Denmark, with its relatively small population, is known for its many validated and complete national registries. With integration of data from health care registries, eg, the Causes of Death Registry and the Danish Cancer Registry,<sup>6,7</sup> we created a new registry (the Danish Prostate Cancer Registry [DaPCaR]), which holds the potential to be able to answer a number of important research questions about epidemiology and pathology of PCa.

## Materials and methods

### Contributing health registries in Denmark

#### The Danish Pathology Register (Patobank)

Patobank is a national nonprofit database that was initiated in 1999 as a registered work tool for Danish pathologists. It holds histopathological information on specimens that have undergone pathological assessment in any Danish hospital or private pathology practice. It is updated continuously and registration of all pathological specimens is mandatory. Since the initiation, data on diagnosis have been backlogged and are complete from 1990 onward, except for one department of pathology that registered data from 1999.<sup>5,8</sup>

According to the bylaws of Patobank, data are available for research purposes. Applications for extraction of data will be considered after approval from the Danish National Data Protection Agency. Approval from the Danish National Data Protection Agency (file number: 2012-41-0390) was obtained.

Data in Patobank are stored using the complex format of SNOMED codes. Each SNOMED code consists of one

letter and five digits, and each code represents information on type of organ, anatomic orientation, diagnosis, grading, histopathological techniques applied, or procedure from where the tissue originated.

A data set with any combination of SNOMED codes including prostate tissue and/or seminal vesicles evaluated in the period 1995–2011 was retrieved in April 2012. The extracted data included the Central Person Registry (CPR) number, type of specimen, involved clinical department, the evaluating department of pathology, specimen number, date of assessment/procedure, patient age at assessment, and SNOMED codes.

During the following, and if not stated otherwise, the term specimen refers to all part elements (eg, core biopsies) in an assembled histopathological assessment (ie, a set of biopsies), corresponding to one pathology report.

#### The Central Person Registry

All Danish citizens are registered by a unique ten-digit CPR number. The first six digits represent the date of birth, while the last four digits provide a unique identification number and information about sex. Established in 1968, the registry includes parameters such as place of birth, place of residence, citizenship, parents' identity, etc.<sup>9–11</sup> CPR is used in all national registries, and the system makes it possible to track an individual in all the available registries.

#### The Causes of Death Registry

Initiated in 1875 and fully computerized in 1970, the Causes of Death Registry holds information on time of death, type of death (ie, natural death, suicide, homicide, and accident), and cause of death. The physician who declares a person dead is obligated to fill in the death certificate, registering both the diagnosis associated directly with death, as well as the underlying cause and contributing diagnoses associated with the cause of death. In cases of autopsies, the referent pathologist is responsible for the death certificate report.<sup>6</sup> Since 2002, all death certificates have been automatically validated through the IRIS – language independent coding system.<sup>12</sup>

#### The Danish Cancer Registry

The Danish Cancer Registry was established in 1947 and reporting has been obligatory since 1987. After a revision in 2008, the Danish Cancer Registry holds data on personal characteristics (CPR number and date of diagnosis) as well as tumor characteristics at the time of diagnosis. Prior to 2004, PCa tumor characteristics were registered according to a simplified WHO grading system, and since 2004, tumor-node-metastasis (TNM) classification (according to International Classification of Diseases [ICD]-10) has been used.<sup>7</sup>

## The Danish Prostate Cancer Registry

The DaPCaR database has been approved by the Danish National Data Protection Agency (file number: 2012-41-0390) and by the science ethics committee of The Capital Region of Denmark (protocol number: H-4-2014-FSP). An overview of DaPCaR is demonstrated in Figure 1.

The DaPCaR database as well as the use of data from the Central Persons Registry, the Cause of Death Registry, Patobank, the Danish Cancer Registry and biochemical departments for the present study was approved by the Danish National Data Protection Agency (file number: 2012-41-0390) and by the ethics committee of The Capital Region of Denmark (protocol number: H-4-2014-FSP). In the approval from the ethics committee, it was stated that due to the nature of the present study, informed consent from patients was not necessary.

All 31 pathological departments contributing with histopathological examinations of prostate tissue in Denmark have complete backlogged data from before 1995, except for one that only had full registration implemented from January 1, 1999. This particular department assessed 1%–2% of the total amount of specimens in DaPCaR from 1995 to 1999. Thus, DaPCaR holds histopathological information on at least 98% of examined prostate tissue in Denmark from 1995 to 2011.

Data were delivered as a comma-separated file, and each line in the file represented one pathology report, corresponding to one histopathological case.

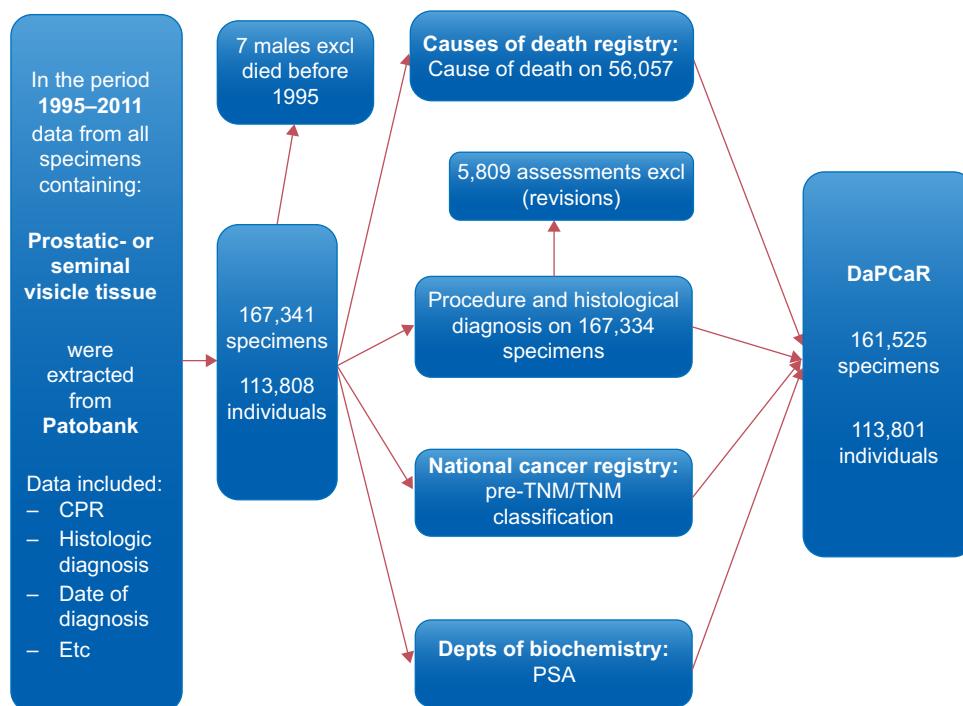
In every line, the evaluating pathologist's assessment report was coded with up to 69 combinations of 1–60 individual SNOMED codes separated by semicolons. A total of 26,295 of these SNOMED code combinations were identified. The code combinations between each semicolon represented one specimen in a case, eg, one prostate core biopsy in a set of biopsies. Examples of typical lines are demonstrated in Figure 2.

The transcoded data extracted from Patobank formed the basis of DaPCaR, and data from other registries was added to the histopathological data based on CPR. Table 1 gives an overview of the contributing health registries and their data contribution.

## Methodology – transcoding of SNOMED codes

To transcode SNOMED codes into a meaningful and analyzable format, a number of computational steps were performed. Figure 3 provides an overview of these steps.

First, each unique SNOMED code combination was translated into text strings according to the official SNOMED<sup>13</sup> classification as of April 2014. Second, these



**Figure 1** Flowchart showing the integration process of data from multiple national registries into the DaPCaR.

**Note:** Revisions: reassessments of specimens, ie, when a patient is referred to another department.

**Abbreviations:** CPR, Central Person Registry; Depts, departments; excl, excluded; PSA, prostate-specific antigen; DaPCaR, Danish Prostate Cancer Registry; TNM, tumor-node-metastasis.

- a) P30990 T77100 M72000 – (bN)
- b) T77100 M81403 P30990 – (bAC)
- c) P30990 T77100 M81403 AEF010A; P30990 T77100 M81403 AEF010A – (bGS5+5)
- d) T77100 M09450 P30990 P3B000; T77100 M09450 P30990 P3B000, T77100 M09450 P30990; T77100 M09450 P30990 P3B000; T77100 M09450 P30990 P3B000; T77100 M81403 AYYYY00 P30990 P3B000; T77100 MB1403 AEF034A P30990 P3B000; T77100 M09450 P30990 P3B000; T77100 M09450 P30990 P3B000; T77100 M09450 P30990 – (bGS3+4)
- e) T08000 M09450 P30841; T08000 M09450 P30841; T77000 M81403 AEF007A AEF034A AEF1843 M09401 P306X0 P30841 – (rapR+GS3+4T2cN0Mx)

**Figure 2** Examples of typical lines of SNOMED code combinations (final conclusion in DaPCaR).

**Notes:** **a)** Simple code consisting of one combination of three SNOMED codes (biopsy of the prostate containing nonmalignant tissue); **b)** simple code with missing information about GS, consisting of one combination of three SNOMED codes (biopsy of the prostate containing adenocarcinoma); **c)** simple code consisting of two identical combinations of three SNOMED codes (biopsy of the prostate containing adenocarcinoma, GS 5+5); **d)** complex code consisting of ten combinations of 3–5 SNOMED codes (biopsy of the prostate containing adenocarcinoma, GS 3+4); **e)** complex code consisting of three or eight combinations of three or eight SNOMED codes (radical prostatectomy with positive surgical margins containing a pT2c adenocarcinoma, GS 3+4 with negative lymph nodes).

**Abbreviations:** bN, biopsy of the prostate with non-malignant tissue; bAC, biopsy of the prostate containing adenocarcinoma; rap, radical prostatectomy; N0Mx, negative lymph nodes and unknown distant metastatic stage; SNOMED, Systematized Nomenclature of Medicine; DaPCaR, Danish Prostate Cancer Registry; GS, Gleason score.

**Table 1** Contributing health registries

Registry	Data contribution
National Pathology Register <sup>5,6</sup> (Patobank)	Histopathologic data Date of diagnosis
Central Person's Registry <sup>7-9</sup>	Vital status (ie, alive, dead, migrated, lost to follow-up)
Causes of Death Registry <sup>10</sup>	Causes of death (PCa, other cause)
National Cancer Registry <sup>12</sup>	Diagnostic tumor classification (ie, TNM classification)
Depts of biochemistry	PSA and PSA ratio Testosterone

**Abbreviations:** Depts, departments; PCa, prostate cancer; PSA, prostate-specific antigen; TNM, tumor-node-metastasis.

text codes were recoded into codes including procedure type, tissue type, and diagnosis, eg, bGS 3+4 (biopsy of the prostate containing a Gleason score [GS] 3+4 adenocarcinoma) or bN (biopsy of the prostate without malignancy). Approximately, half of the examinations contained multiple SNOMED code combinations, eg, two normal core biopsies and several core biopsies with PCa. To reach a final conclusion for every unique histopathological assessment in the Patobank data set, a third computational step was therefore included to prioritize and summarize conclusions for each specimen.

The Gleason system grades cell differentiation on a scale from one to five, where 1 represents “highly differentiated cells” and 5 represents the “most dedifferentiated cells”. The final GS for a particular PCa is the sum of the Gleason grades, representing the worst degree of differentiation and the most frequent Gleason pattern observed. All institutions apart from one reported GS as one score summarizing the total number of core biopsies in a biopsy set. The GS reported from these institutions was accepted as the final conclusion. In the present data set, one institution registered one GS per

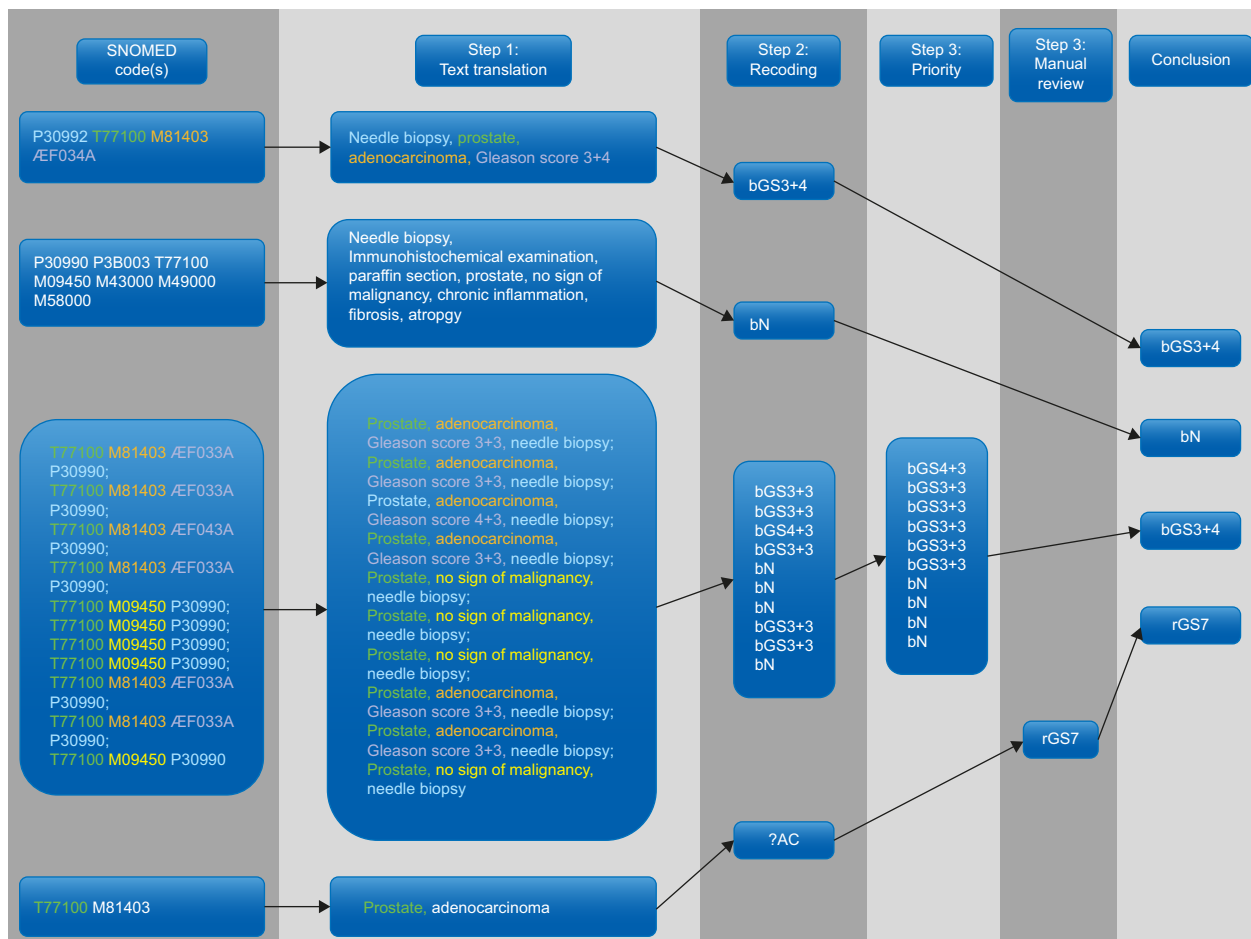
core biopsy, as opposed to one GS per biopsy set. In 873 (1% of all prostate biopsies) of these cases, a set of biopsies had been given multiple GSs. To make results comparable between institutions and to avoid upgrading and reporting of multiple GSs per patient, these 873 specimens were reviewed and divided into three groups (GS ≤6, GS =7, and GS ≥8).

Almost 45,000 assessments had insufficient coding, ie, no specified GS or missing information about procedure type. To overcome this, the text part of the histopathological assessment rapport of these assessments was reviewed to reach a final conclusion (computational step 4).

Many males had multiple procedures performed, ie, rebiopsies on active surveillance or diagnostic core biopsies followed by total prostatectomy, and were therefore represented in the database by multiple PCa-containing specimens during the inclusion period. The term “diagnostic procedure” was defined as the first procedure resulting in a specimen containing PCa in each individual patient case.

A total of 5,809 specimens had undergone pathological reassessment, typically in relation to referral to a tertiary treating facility prior to surgery. Reassessment is coded by a specific code in the SNOMED system. To ensure that each specimen was represented only once in DaPCaR, all examination reports from patients having at least one reassessment were manually reviewed. If the conclusions of the reassessment differed from the original conclusion, the reassessment conclusion was used.

Procedures were separated into distinguished categories as mentioned in Table 2. Transrectal ultrasound (TRUS)-guided core biopsies of the prostate were distinguished from TRUS-guided core biopsies of the seminal vesicles, anastomoses, or periprostatic area only. If TRUS-guided core



**Figure 3** Computational steps in the process of transcoding SNOMED codes into a readily analyzable format.

**Abbreviations:** b, biopsy of the prostate; GS, Gleason score; N, nonmalignant; ?, unknown procedure; AC, adenocarcinoma; r, transurethral resection of the prostate; SNOMED, Systematized Nomenclature of Medicine.

biopsies of the prostate were obtained in the same specimen (ie, same assessment) as TRUS-guided core biopsies of the periprostatic area/seminal vesicles/anastomotic area and/or transurethral resection of the prostate (TUR-P), the set was collectively regarded as “TRUS-guided core biopsies of the prostate”. Prostate tissue removed by transvesical approach was, together with TUR-Ps, regarded as “resections of the prostate”. “Radical prostatectomies” included prostates removed by open retroperitoneal approach or by laparoscopy (+/- robot assisted). Transurethral biopsies of the bladder (TUR-Bs) containing benign prostate tissue were distinguished from TUR-Bs containing prostate cancer and regarded as either “resection of prostate + bladder” or “resection of the prostate”. Cases in which it could not be determined whether the procedure was a bladder only or a bladder and prostate intervention were regarded as “prostate + bladder resection”. Total cystoprostatectomy specimens, specimens from gastrointestinal interventions, autopsies, and specimens from bladder biopsies/bladder resections were translated into specific conclusions. Specimens with

inconclusive diagnoses and specimens from female anatomical sites and otherwise meaningless coding were defined as “not suitable”.

All nonmalignant diagnoses were regarded as normal, including diagnoses such as inflammation and granulomas. All adenocarcinomas coded by the WHO classification as having high, intermediate, or low differentiation or as being a grade 1, 2, or 3 adenocarcinoma were in both circumstances translated into  $GS \leq 6$ ,  $GS = 7$ , and  $GS \geq 8$ , respectively.

Locally advanced disease and metastatic disease were defined as PCa cells invading the periprostatic tissue and PCa cells within any defined organ (except from the bladder) other than the prostate, respectively. Neuroendocrine differentiation and small cell carcinoma components were registered specifically, also in cases where a specimen was also GS graded.

If the diagnosis was a possible prostate/urothelial cancer (ie, uncertainty as to whether invasive or not) or certain dysplasia, it was registered specifically as possible adeno-

**Table 2** Translation of SNOMED codes into conclusions

SNOMED code	Conclusion
TRUS-guided biopsy of the prostate	Biopsy of the prostate
TRUS-guided biopsies of the prostate + periprostatic area and/or seminal vesicles and/or anastomoses and/or transurethral resection of the prostate	
TRUS-guided biopsy of the periprostatic area only	Biopsy of the periprostatic area only
TRUS-guided biopsy of the seminal vesicle only	Biopsy of the seminal vesicle only
TRUS-guided biopsy of the anastomoses only	Biopsy of the anastomoses only
Transurethral resection of the prostate	Resection of the prostate
Transurethral resection of bladder and prostate with prostate cancer	
Transvesical resection of the prostate	
Transurethral resection of prostate and bladder without prostate cancer or with bladder cancer and where the indication for intervention was unknown	Resection of the prostate+bladder
Resection of seminal vesicles only	Resection of seminal vesicles only
Resection of anastomoses only	Resection of anastomoses only
Aspirations or cytological (asp/cyt) examination of the prostate	Asp/cyt of the prostate
Asp/cyt examination of the seminal vesicles only	Asp/cyt of the seminal vesicles only
Procedure not specified from the prostate	Unknown prostatic procedure
Procedure not specified from the seminal vesicles only	Unknown seminal vesicle procedure
Open retroperitoneal removal of the prostate	Radical prostatectomy
Laparoscopic (+/- robot assisted) removal of the prostate	
Cystectomy	Radical cystoprostatectomy
Cystectomy + prostatectomy	
Autopsies	Autopsies
Transurethral biopsy of the bladder	Bladder biopsies/resection
Transurethral resection of the bladder	
Gastrointestinal (GI) interventions	GI-interventions
Meaningless or otherwise inconclusive coding	Not suitable

**Notes:** Blue, biopsies; red, resections; green, aspirations; orange, unknown procedures; gray: other procedures.

**Abbreviations:** TRUS, transrectal ultrasound; SNOMED, Systematized Nomenclature of Medicine.

carcinoma or prostatic intraepithelial neoplasia, respectively. Possible dysplasia was registered as “nonmalignant”.

In addition, all histopathological assessment reports from radical prostatectomies were reviewed manually to ensure complete data regarding surgical margin status, GS, and pTNM stage to be further analyzed.

The computational algorithm for transcoding SNOMED codes into analyzable conclusions was developed in the .xlsx format by our data manager (Günther Momsen) using Microsoft Excel 2010. Statistical analysis was performed using SPSS statistics version 22 (IBM Corporation, Armonk, NY, USA).

## Results

A total of 167,341 specimens from 113,808 males were extracted from Patobank. Seven males died before 1995 and were excluded. A total of 5,809 specimens were revised at a tertiary pathology department, typically before curative-intended therapy. We decided that the final conclusion was retrieved from the revision, and hence the original report was excluded. Therefore, 161,525 unique histopathological assessments from 113,801 individuals were included in DaPCaR. Of these, a histopathological diagnosis of PCa was recorded

**Table 3** Diagnostic procedures

Diagnosis and diagnostic procedure	Individuals	Percentage
Prostate cancer	48,078	100
Biopsy of the prostate	37,895	78.82
Vesicular biopsy only	9	0.019
Periprostatic biopsy only	1	0.0021
TUR/transvesical resection of the prostate	8,061	16.77
Aspiration of the prostate	33	0.068
Vesicular resection only	1	0.0021
Radical prostatectomies	75	0.16
Radical cystoprostatectomies	488	1.01
Autopsies	749	1.56
Gastrointestinal interventions	9	0.021
Unknown	757	1.57

**Abbreviation:** TUR: transurethral resection.

in 48,078 (42.2%) males. Table 3 shows an overview of how these males were diagnosed, defined as the first procedure resulting in a specimen containing PCa. Of these PCa cases, 78.8% and 16.8% were diagnosed by TRUS-guided core biopsies and TUR-Ps, respectively. In only 1.5% of the cases, the diagnostic procedure was unknown. The remaining 3.5% were according to Table 3 diagnosed otherwise (ie, unexpected during radical cystoprostatectomies or autopsies

or during assessment of seminal vesicle biopsy as a staging procedure for biochemically or clinically diagnosed PCa, etc).

A complete overview of final histopathological conclusions is displayed in Table 4. A total of 85,238 specimens (52.8% of all specimens) were from TRUS-guided core biopsies. Of these, 97.8% were from core biopsies of the prostate, and ~2% of TRUS-guided core biopsies originated from periprostatic area, seminal vesicles, or anastomoses. A total of 59,726 cases (37% of all specimens) derived from resections. And of these, 94% were from TUR-Ps or transvesical resections of the prostate, and 6% were from combined prostate and bladder resections. A total of 129 (0.08% of all specimens) were aspiration biopsies, ie, cytology specimens, and of these, 90.7% were from the prostate. In 3,382 cases (2.1% of all specimens), the procedure could not be specified, but 99.9% of these derived from the prostate. Overall, 6,848 (4.24%) cases were from radical prostatectomies, 2,506 (1.55%) from autopsies, 1,451 (0.89%) from radical cystoprostatectomies, 1,430 (0.89%) from bladder interventions, and 232 (0.14%) from gastrointestinal interventions. Only in 583 (0.36%) cases, the coding for topography was meaningless.

PCa was identified in 50% of all TRUS-guided prostate biopsy sets, in 9.7% of TRUS-guided biopsies of seminal vesicle, in 21.7% of all TUR-Ps/transvesical resections of the prostate, and in 20.8% of all aspiration biopsies. In specimens from unknown prostatic procedures, PCa was identified in 27.8%.

## Discussion and perspectives

Currently, there is no national comprehensive Danish database on PCa. With several national registries available in Denmark, we initiated the DaPCaR program to create a novel national registry of Danish males with PCa, containing complete histopathological data for all males who have undergone evaluation of prostate tissue in Denmark during a 16-year period. Using a newly developed algorithm, we succeeded in converting complex SNOMED code combinations into an analyzable format suitable for PCa epidemiology research. With DaPCaR, we have a flexible platform for integration with other registries. An interesting attribution of DaPCaR is the fact that it does not only hold information on males with PCa but also holds information on autopsies, cystoprostatectomies, and data on males with negative TRUS biopsies who underwent evaluation for PCa, which is, to the best of our knowledge, unprecedented.<sup>14-17</sup> DaPCaR offers a complete overview of the histopathological diagnostic activity within prostatic diseases during a 16-year period.

The most important caveat in the creation of this data platform is internal and external validity of the transcoded SNOMED codes. Several validation efforts were made. Every unique SNOMED code has been manually translated into text strings to ensure that all clinically relevant information in the SNOMED codes of Patobank was preserved. The use of computer programming ensured that identical SNOMED code combinations were translated into identical text strings and later into identical diagnoses and conclusions in DaPCaR. The prioritizing and summarizing of conclusions in cases of multiple SNOMED codes (translational step 3) were done manually to ensure that no information was lost. In order to reduce the number of specimens with missing information, more than 55,000 pathology reports were reviewed manually (6,892 radical prostatectomies, 5,809 reassessed specimens, and ~45,000 additional specimens).

An external validation of DaPCaR was made through other registries and publications. A recent publication on radical prostatectomies performed in Denmark based on local registries identified 6,489 procedures performed in the period 1995–2011.<sup>18</sup> All these patients were recognized in DaPCaR, and in addition, 359 males with PCa who underwent radical prostatectomy (RP) in the period were identified in DaPCaR.

A total of 75 patients had an RP performed with no prior biopsy procedure recorded. These patients were either biopsied before 1995, diagnosed on core biopsies abroad, or biopsied in private practice, from where Patobank registration was not mandatory until 2005.

Additional external validation was performed using The Nordic Cancer Registry (NORDCAN),<sup>1</sup> that holds information on incidence, prevalence, and mortality among 50 cancer types in the northern countries.<sup>1</sup> The Danish contribution to NORDCAN is achieved through integration of data from the Causes of Death Registry and the Danish Cancer Registry. In DaPCaR, we identified 48,078 patients with histopathologically verified PCa, and in the same period, NORDCAN registered 49,863 cases of PCa (difference of 3.6%). The obvious explanation for this difference is that some PCAs are reported based on clinical findings without histological confirmation (eg, a man with palpable PCa on digital rectal examination, significantly elevated PSA, and metastasis verified by bone scan). This is supported in a recent publication showing that in 1997, 13% of all diagnosed PCAs in Denmark were not histopathologically verified.<sup>19</sup> We believe that the narrow margin between DaPCaR and NORDCAN suggests a high validity of our transcoded data set. Differences in the number of PCa cases between DaPCaR and NORDCAN decreased over time, suggesting a higher prevalence of

**Table 4** Overview of specimens and diagnosis

Specimen type, location, and overall diagnosis	Specimens	Percentage
<b>Biopsies</b>	85,238	52.77 <sup>a</sup>
Prostate	83,379	97.82 <sup>b</sup>
Prostate cancer	41,674	49.98 <sup>c</sup>
Possible prostate cancer	1,813	2.17 <sup>c</sup>
Dysplasia (prostatic)	1,662	1.99 <sup>c</sup>
Bladder cancer	97	0.12 <sup>c</sup>
Possible bladder cancer	12	0.014 <sup>c</sup>
Normal	38,065	45.65 <sup>c</sup>
Other	56	0.067 <sup>c</sup>
Periprostatic only	18	0.021 <sup>b</sup>
Prostate cancer	6	33.33 <sup>d</sup>
Normal	12	66.67 <sup>d</sup>
Seminal vesicular only	1,831	2.15 <sup>b</sup>
Prostate cancer	177	9.67 <sup>e</sup>
Possible prostate cancer	4	0.22 <sup>e</sup>
Bladder cancer	2	0.11 <sup>e</sup>
Possible bladder cancer	1	0.054 <sup>e</sup>
Normal	1,644	89.79 <sup>e</sup>
Other	3	0.16 <sup>e</sup>
Anastomotic only	10	0.012 <sup>b</sup>
Prostate cancer	7	70.0 <sup>f</sup>
Normal	3	30.0 <sup>f</sup>
<b>Resections (TUR/transvesical)</b>	59,726	36.98 <sup>a</sup>
Prostate	56,118	93.96 <sup>g</sup>
Prostate cancer	12,167	21.68 <sup>h</sup>
Possible prostate cancer	171	0.30 <sup>h</sup>
Dysplasia (prostatic)	186	0.33 <sup>h</sup>
Bladder cancer	442	0.79 <sup>h</sup>
Possible bladder cancer	11	0.020 <sup>h</sup>
Normal	43,013	76.65 <sup>h</sup>
Other	128	0.23 <sup>h</sup>
Prostate+bladder	3,589	6.01 <sup>g</sup>
Bladder cancer	1,847	51.46 <sup>i</sup>
Possible bladder cancer	38	1.06 <sup>i</sup>
Urothelial dysplasia	6	0.17 <sup>i</sup>
Normal	1,676	46.70 <sup>i</sup>
Other	22	0.61 <sup>i</sup>
Seminal vesicular only	17	0.028 <sup>g</sup>
Prostate cancer	2	11.76 <sup>j</sup>
Bladder cancer	2	11.76 <sup>j</sup>
Normal	11	64.71 <sup>j</sup>
Other	2	11.76 <sup>j</sup>
Anastomotic only	2	0.0033 <sup>g</sup>
Normal	2	100 <sup>k</sup>
<b>Aspirations and cytology</b>	129	0.080 <sup>a</sup>
Prostate	117	90.70 <sup>l</sup>
Prostate cancer	33	28.21 <sup>m</sup>
Possible prostate cancer	4	3.42 <sup>m</sup>
Bladder cancer	1	0.85 <sup>m</sup>
Normal	72	61.54 <sup>m</sup>
Other	7	5.98 <sup>m</sup>
Seminal vesicular only	12	9.30 <sup>l</sup>
Normal	12	100 <sup>n</sup>
<b>Procedure not specified</b>	3,382	2.09 <sup>a</sup>
Prostate	3,378	99.88 <sup>o</sup>
Prostate cancer	939	27.80 <sup>p</sup>
Possible prostate cancer	21	0.62 <sup>p</sup>

(Continued)

**Table 4** (Continued)

Specimen type, location, and overall diagnosis	Specimens	Percentage
Dysplasia (prostatic)	7	0.21 <sup>p</sup>
Bladder cancer	12	0.36 <sup>p</sup>
Normal	2,390	70.75 <sup>p</sup>
Other	9	0.27 <sup>p</sup>
Seminal vesicular only	4	0.12 <sup>o</sup>
Normal	3	75.0 <sup>q</sup>
Other	1	25.0 <sup>q</sup>
<b>Radical prostatectomies</b>	6,848	4.24 <sup>a</sup>
<b>Radical cystoprostatectomies</b>	1,451	0.89 <sup>a</sup>
<b>Autopsies</b>	2,506	1.55 <sup>a</sup>
<b>Bladder biopsies/resection</b>	1,430	0.89 <sup>a</sup>
<b>Gastrointestinal interventions</b>	232	0.14 <sup>a</sup>
<b>Not suitable</b>	583	0.36 <sup>a</sup>
<b>Total</b>	<b>161,525</b>	<b>100</b>

**Notes:** <sup>a</sup>Percentage of specimens; <sup>b</sup>percentage of TRUS-guided biopsies; <sup>c</sup>percentage of prostate biopsies; <sup>d</sup>percentage of periprostatic biopsies; <sup>e</sup>percentage of seminal vesicle biopsies; <sup>f</sup>percentage of anastomotic biopsies; <sup>g</sup>percentage of resections; <sup>h</sup>percentage of prostate resections; <sup>i</sup>percentage of prostate+bladder resections; <sup>j</sup>percentage of seminal vesicle resections; <sup>k</sup>percentage of anastomotic resections; <sup>l</sup>percentage of aspirations/cytology; <sup>m</sup>percentage of aspirations/cytology of the prostate; <sup>n</sup>percentage of aspirations/cytology of the seminal vesicle; <sup>o</sup>percentage of unknown procedures; <sup>p</sup>percentage of unknown prostatic procedures; <sup>q</sup>percentage of unknown seminal vesicle procedures. Blue, biopsies; red, resections; green, aspirations; orange, unknown procedures; gray: other procedures.

**Abbreviations:** TUR, transurethral resection; TRUS, transrectal ultrasound.

histologically verified PCa today compared to the beginning of the investigated period.

With transcoding and integration of Patobank data into DaPCaR, a large and complete national database containing validated histological data on males who have undergone clinical and histopathological evaluation of the prostate has been created. Further, clinically relevant data have been integrated from other national registries using the unique Danish CPR number system.<sup>9-11</sup>

From the Danish Cancer Registry, diagnostic tumor classification has been integrated into DaPCaR. Prior to 2004, a simplified WHO classification system defined as pre-TNM was used to classify tumors, and since 2004, tumors have been classified according to the TNM system.<sup>7</sup> In DaPCaR, data regarding tumor classification follow two different classification systems, pre-TNM for patient diagnosed prior to 2004 and TNM for patients diagnosed after 2004. The Danish Cancer Registry and its information has previously been validated in PCa patients.<sup>19</sup>

Vital status was obtained from the Central Person Registry, and for the proportion of males who had died on April 28, 2015, the cause of death was obtained from the Causes of Death Registry. This registry holds information on time of death, overall type of death (ie, natural death, suicide, homicide, accident), the condition directly causing death, as well as underlying cause and contributory causes of death.<sup>6</sup>



Serum PSA, PSA ratio (ie, serum-PSA/free-PSA level), and *s*-testosterone values for all patients in DaPCaR were collected from local databases and regional laboratory databases. Unfortunately, several of those databases do not hold information on all males, especially in the early period of investigation, and thus only 47.4% of males with PCa had diagnostic laboratory data recorded.

The Danish CPR system provides the opportunity to integrate information from a range of national registries including income information, geographical distribution, and other social parameters as well as information regarding received PCa treatment and complications. The National Patient Registry (Landspatientregistret)<sup>20</sup> holds data on every hospitalization and outpatient treatment in Denmark. Data are stored according to CPR and ICD codes for diagnosis and procedures, which make it possible to track the type of treatment given to each patient as well as whether the patient had complications necessitating hospitalization or outpatient visits.<sup>20,21</sup> This holds the potential for DaPCaR to be even more comprehensive and thus an even more promising platform for future studies of PCa.

At present, DaPCaR is a closed database; however, it is the intention to maintain the database, add new cases, and increase the length of follow-up through future updates.

## Conclusion

By decrypting complex coding from Patobank into analyzable conclusions, we managed to create a complete national Danish histological registry on males undergoing histopathological examination of prostate tissue in the period 1995–2011. DaPCaR was internally and externally validated through several computational steps and comparison to other registries. Integration of clinically relevant information from other high-quality national registries makes DaPCaR a novel platform for important future studies of PCa in Denmark.

## Acknowledgments

The authors would like to acknowledge the information technology assistance of data manager Günther Momsen. The creation of the DaPCaR would not have been possible without his expert Microsoft Excel skills. The study was funded by grants from IMK (Ib Mogens Kristiansen) general funds, the Gangsted fund, The Ing-Britt and Stig Mårtensson Foundation, and the Capital Region of Denmark.

## Disclosure

The authors report no conflicts of interest in this work.

## References

1. Engholm G, Ferlay J, Christensen N, et al. *NORDCAN: Cancer Incidence, Mortality, Prevalence and Survival in the Nordic Countries, Version 7.1 (09.07.2015)*. Association of the Nordic Cancer Registries. Danish Cancer Society; 2015. Available from: <http://www.ancr.nu>. Accessed July 9, 2015.
2. Center MM, Jemal A, Lortet-Tieulent J, et al. International variation in prostate cancer incidence and mortality rates. *Eur Urol*. 2012;61(6):1079–1092.
3. Mukai TO, Bro F, Pedersen KV, Vedsted P. [Use of prostate-specific antigen testing]. *Ugeskr Laeger*. 2010;172(9):696–700. Danish.
4. Brasso K, Ingimarsdóttir IJ, Rusch E, et al. Differences in survival from prostate cancer in Denmark, Iceland and Sweden. *Eur J Cancer*. 2013;49(8):1984–1992.
5. Bjerregaard B, Larsen OB. The Danish Pathology Register. *Scand J Public Health*. 2011;39(Suppl 7):72–74.
6. Helweg-Larsen K. The Danish Register of Causes of Death. *Scand J Public Health*. 2011;39(Suppl 7):26–29.
7. Gjerstorff ML. The Danish Cancer Registry. *Scand J Public Health*. 2011;39(Suppl 7):42–45.
8. Patobanken, Sundhedsstyrelsen. *Fællesindhold for Basisregistrering Af Patologisk-Anatomiske Undersøgelser* [Common Contents for Basic Registration of Pathological-Anatomical Studies]; 2011. Danish.
9. Nielsen H. *CPR – Danmarks Folkeregister* [CPR – Denmark's National Register]. CPR kontoret; 1991.
10. Pedersen CB, Gøtzsche H, Møller JO, Mortensen PB. The Danish Civil Registration System. A cohort of eight million persons. *Dan Med Bull*. 2006;53(4):441–449.
11. Pedersen CB. The Danish Civil Registration System. *Scand J Public Health*. 2011;39(Suppl 7):22–25.
12. Johansson LA, Pavillon G. *IRIS: A Language-Independent Coding System Based on the NCHS System MMDS*; WHO-FIC network meeting, Tokyo, Japan, 2005. Available from: [http://apps.who.int/classifications/apps/icd/meetings/tokyomeeting/B\\_6-2\\_IRIS\\_A\\_language\\_independent\\_coding\\_system.pdf](http://apps.who.int/classifications/apps/icd/meetings/tokyomeeting/B_6-2_IRIS_A_language_independent_coding_system.pdf). Accessed May 15, 2015.
13. Côté RA, Robboy S. Progress in medical information management. Systematized nomenclature of medicine (SNOMED). *JAMA*. 243(8).
14. Scosyrev E, Messing J, Noyes K, Veazie P, Messing E. Surveillance Epidemiology and End Results (SEER) program and population-based research in urologic oncology: an overview. *Urol Oncol*. 2012;30(2):126–132.
15. Hågel E, Garmo H, Bill-Axelsson A, et al. PCBaSe Sweden: a register-based resource for prostate cancer research. *Scand J Urol Nephrol*. 2009;43(5):342–349.
16. Van Hemelrijck M, Garmo H, Wigertz A, Nilsson P, Stattin P. Cohort Profile Update: The National Prostate Cancer Register of Sweden and Prostate Cancer data Base – a refined prostate cancer trajectory. *Int J Epidemiol*. 2016;45(1):73–82.
17. Cooperberg MR, Broering JM, Litwin MS, et al. The contemporary management of prostate cancer in the United States: lessons from the cancer of the prostate strategic urologic research endeavor (CapSURE), a national disease registry. *J Urol*. 2004;171(4):1393–1401.
18. Røder MA, Brasso K, Christensen IJ, et al. Survival after radical prostatectomy for clinically localised prostate cancer: a population-based study. *BJU Int*. 2014;113(4):541–547.
19. Ingimarsdóttir IJ, Rusch E, Engholm G, Storm HH, Brasso K. Quality assessment of prostate cancer reports to the Danish Cancer Registry. *Acta Oncol*. 2016;55(1):24–29.
20. Lyng E, Sandegaard JL, Rebolj M. The Danish National Patient Register. *Scand J Public Health*. 2011;39(Suppl 7):30–33.
21. Mason K, Thygesen LC, Stenager E, Brønnum-Hansen H, Koch-Henriksen N. Evaluating the use and limitations of the Danish National Patient Register in register-based research using an example of multiple sclerosis. *Acta Neurol Scand*. 2012;125(3):213–217.

## Clinical Epidemiology

Dovepress

### Publish your work in this journal

Clinical Epidemiology is an international, peer-reviewed, open access, online journal focusing on disease and drug epidemiology, identification of risk factors and screening procedures to develop optimal preventative initiatives and programs. Specific topics include: diagnosis, prognosis, treatment, screening, prevention, risk factor modification,

systematic reviews, risk and safety of medical interventions, epidemiology and biostatistical methods, and evaluation of guidelines, translational medicine, health policies and economic evaluations. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use.

Submit your manuscript here: <https://www.dovepress.com/clinical-epidemiology-journal>