

# Existing data sources for clinical epidemiology: the Danish National Pathology Registry and Data Bank

Rune Erichsen<sup>1</sup>  
Timothy L Lash<sup>1</sup>  
Stephen J Hamilton-Dutoit<sup>2</sup>  
Beth Bjerregaard<sup>3</sup>  
Mogens Vyberg<sup>2,4</sup>  
Lars Pedersen<sup>1</sup>

<sup>1</sup>Department of Clinical Epidemiology, <sup>2</sup>Institute of Pathology, <sup>4</sup>Aalborg Hospital, Aarhus University Hospital, Aarhus, Denmark; <sup>3</sup>Department of Pathology and the Pathology Data Bank, Herlev University Hospital, Herlev, Denmark

**Abstract:** Diagnostic histological and cytological specimens are routinely stored in pathology department archives. These biobanks are a valuable research resource for many diseases, particularly if they can be linked to high quality population-based health registries, allowing large retrospective epidemiological studies to be carried out. Such studies are of significant importance, for example in the search for novel prognostic and predictive biomarkers in the era of personalized medicine. Denmark has a wealth of highly-regarded population-based registries that are ideally suited to conduct this type of epidemiological research. We describe two recent additions to these databases: the Danish National Pathology Registry (DNPR) and its underlying national online registration database, the Danish Pathology Data Bank (DPDB). The DNPR and the DPDB contain detailed nationwide records of all pathology specimens analyzed in Denmark since 1997, and an incomplete but nonetheless valuable record of specimens from some pathology departments dating back to the 1970s. The data are of high quality and completeness and are sufficient to allow precise and efficient localization of the specimens. We describe the relatively uncomplicated procedures required to use these pathology databases in clinical research and to gain access to the archived specimens.

**Keywords:** cohort profile, epidemiology, paraffin blocks, pathology

Each year, many thousands of diagnostic histological and cytological specimens pass through pathology laboratories throughout the world. Having been used to establish an immediate clinical diagnosis, these specimens are then stored in the pathology archives. In the case of tissue specimens, the vast majority are fixed in formalin, embedded in paraffin blocks, and are kept indefinitely. These archived paraffin blocks represent by far the largest tissue biobanks in the world, and they are a valuable research resource. This paraffin biobank has become of increasing importance due to technical advances that mean that a wide range of molecular biological techniques can now be applied to DNA, RNA, and proteins extracted from routinely archived paraffin blocks (and in some cases from archived glass slides, including cytological specimens), making possible large scale molecular epidemiological studies on even rare diseases. This biobank can be used for research into many types of disease. However, its value as a resource for identifying novel prognostic and predictive biomarkers in cancer has been recently underlined by the emergence of personalized medicine. Large molecular epidemiological studies using routinely archived pathological specimens can only be performed if the pathology biobank is linked to high quality clinical data, as is now the case in Denmark.

Danish health care, disease, and population registries are recognized as being amongst the best in the world, in particular because of their large size, long

Correspondence: Rune Erichsen  
Department of Clinical Epidemiology,  
Aarhus University Hospital, Olof Palmes  
Allé 43-45, DK-8200 Aarhus N, Denmark  
Tel +45 8942 4895  
Fax +45 8948 4801  
Email re@dce.au.dk

record period, high quality, and completeness.<sup>1,2</sup> The long list of population-based registries in Denmark was added to in 1997 by the establishment of the Danish National Pathology Registry (DNPR), facilitating the use of nationwide pathology information in clinical research. Wide-spread health registration in Denmark is based on the Danish National Health Service, which is coordinated through five administrative regions. This health system provides what is essentially a monopoly service of free (tax-funded) medical care for all 5.4 million Danish residents, as well as partially reimbursing costs of most physician-prescribed drugs. With regard to pathology services, there are presently 19 hospital-based pathology departments that handle virtually all pathology specimens examined in Denmark. Not only do all Danish population-based registries have a high degree of completeness, they can also be efficiently linked by a unique personal 10-digit identifier, known as a CPR number. Since 1968, CPR numbers have been assigned to all Danish citizens and residents at birth or immigration through the Civil Registration System. In this paper, we will present an overview of the DNPR and its underlying national registration database, called the Danish Pathology Data Bank (DPDB). We hope this will help researchers to understand the Danish database registry system, and learn some of the potential uses that can be made of Danish pathology data in clinical research.

## The Danish Pathology Data Bank

Computer-based recording of pathological specimens began in Denmark in the 1970s. However, since 1997 all Danish pathology departments have recorded electronically standard data on all biological specimens according to national guidelines for uniform registration published by the Danish National Board of Health (see section “What is recorded?”).<sup>3</sup> This standardization paved the way for the development DPDB, which was established in 1999 as a nationwide database at Hvidovre Hospital in Copenhagen.<sup>4</sup> The DPDB is the first national online health care registry in Denmark. It has a primary function as a routine online tool for diagnostic pathologists throughout Denmark. Once a new pathological specimen is received for a particular patient, it is registered in the system by the examining pathology department, after which information on all previous pathology examinations for that patient can be retrieved online by the pathologist, regardless of where they were previously performed. When the new pathology analysis is complete, data are automatically sent online to update the DPDB. Thus, the DPDB offers instantly updated nationwide information on pathology investigations and diagnoses. Importantly, historical pathology data have been added to a variable extent

to the database. Thus, for some pathology departments, data as far back as the early 1970s have been included (Table 1). In addition to its function as a diagnostic tool, the DPDB is also used to coordinate screening programs and provides an option for pathology staff to assess workload and costs, and to perform quality control assessments.

## The Danish National Pathology Registry

The DNPR was established January 1st, 1997 by the Danish National Board of Health. Reporting to the DNPR has been based on national guidelines for uniform registration<sup>4</sup> and it now takes place automatically through the DPDB. At present, historical information from the period before 1997 is being transferred from the DPDB to the DNPR (Table 1) and, thus, in the near future the DNPR and the DPDB will have the exact same information. The DNPR is funded and managed by the National Board of Health situated in Copenhagen.

The DNPR was established for five major reasons: (1) To serve as a nationwide database of pathology investigations managed by the National Board of Health, a prerequisite for establishing a national computer-based cancer registration scheme. Indeed, there is now a continuous exchange of information between the DNPR and the Danish Cancer Registry to secure accurate and complete cancer registration in Denmark. (2) To improve a number of other registries managed by the National Board of Health (eg, the Danish Registry of Causes of Death, the National Registry of Patients, and the Registry of Congenital Abnormalities) by providing linkage to a national pathology registry. (3) As a tool for managing health resources and evaluating and predicting health-care expenses. (4) As an essential step to providing a national pathology database for medical research. (5) As a tool for maintaining and coordinating various research biobanks.

## What is recorded?

For the DNPR to fulfill the aims described above, all pathology examinations conducted in Denmark have to be registered in a uniform way. Thus, information in the DPDB is recorded in accordance with the national guidelines and there is daily reporting of data from the DPDB to the DNPR, ensuring data uniformity. The following information is reported from the DPDB to the DNPR:<sup>4</sup>

1. The investigating authority: The pathology department or the practicing pathologist (forensic medicine departments do not report to the DPDB or the DNPR because of legal restrictions).
2. Unique ID-number for the requisition/specimen.

**Table 1** Overview of the content of the Danish Pathology Data Bank (<http://www.patobank.dk/>)

Danish regions	Hospital	Since (Date)	What is recorded?
North	Hjørring- Brønderslev	01.09.1968:	All malignant neoplasms
		01.01.1993:	All examinations
North	Aalborg	01.01.1972:	Malignant neoplasms, resections, liver and kidney biopsies
		01.01.1983:	All examinations
Central	Skive	01.01.1972:	Diagnoses only
		01.04.1979:	All examinations
Central	Randers	01.01.1981:	All examinations (with few exceptions)
Central	Aarhus, Tage-Hansens Gade	01.04.1970:	All examinations (with few exceptions)
Central	Aarhus, Nørrebrogade	01.01.1981:	All examinations (with few exceptions)
Central	Holstebro	1980:	Cervix cytology examinations
		01.01.1999:	All examinations
Central	Vejle	01.08.1979:	Cervix cytology
		16.08.1993:	All examinations
South	Esbjerg	01.01.1976:	All examinations
South	Sønderborg	01.03.1975:	5% of examinations, not chronological
		01.01.1999:	All examinations (with exceptions)
South	Odense	01.04.1988:	Diagnoses only
		01.04.1989:	Cervix cytology examinations
		01.01.1996:	All examinations
South	Svendborg	01.04.1988:	Diagnoses only
		01.04.1989:	Cervix cytology examinations
		01.01.1996:	All examinations
Zealand	Roskilde	01.01.1985:	All examinations (no descriptions before 01.01.1990)
Zealand	Nykøbing F	01.01.1977:	Some examinations
		01.01.1982:	All examinations
Zealand	Næstved	01.01.1970:	Some examinations
		01.01.1982:	All examinations
Zealand	Slagelse/Holbæk	01.01.1982:	Diagnoses only
		01.01.1999:	All examinations
Copenhagen	Gentofte	01.01.1989:	All examinations (no descriptions before 01.01.1990)
Copenhagen	Glostrup	01.01.1990:	All examinations
Copenhagen	Herlev	01.01.1985:	All examinations
			(no descriptions before 01.01.1990)
Copenhagen	Hillerød	01.01.1972:	All examinations
Copenhagen	Bispebjerg	01.01.1975:	Some examinations
		01.01.1983:	All examinations
Copenhagen	Frederiksberg	01.01.1981:	Closed 31.12.1996
Copenhagen	Hvidovre	01.01.1976:	All examinations
Copenhagen	Kommune Hospitalet	01.07.1980:	Closed 31.12.1987
Copenhagen	Sundby	01.09.1976:	Closed 31.12.1996
Copenhagen	Rigshospitalet	01.01.1975:	Some examinations
		01.01.1987:	All examinations
			Neuropathology from 01.01.1997 Pediatric pathology from 01.01.1989 Ocular pathology from 01.01.2007

**Notes:** The Danish Pathology Data Bank was established in 1999 but historical data has to some extent been added as indicated in this table. Some of the pathology departments have been merged and at present, 19 pathology departments examine pathology specimens in Denmark. "All examinations" denotes records including gross and microscopy descriptions. "Diagnoses only" denotes records containing no gross or microscopy description.

3. Patient data (CPR number; county of residence; for children, the CPR number of the mother).
4. Requesting hospital department or general practitioner.
5. The hospital department responsible for treatment.
6. Dates of request, reception, and sign out.
7. Specimen type (21 subtypes, eg, histology, cytology, and autopsy).
8. Number of specimens.
9. Procedures performed (eg, paraffin block; tissue slide; immunohistochemical staining; DNA-analysis).
10. Conclusion/free text diagnoses.
11. Diagnoses based on the Danish SNOMED (Systemized Nomenclature of Medicine) codes.
12. Free text for the SNOMED codes.
13. Gross and microscopy descriptions.

Because of the unique ID-number of each pathology examination it is possible to distinguish new records from revisions and, thus, pathologists are able to add new information to the records. Old information cannot be erased, which ensures that an accurate historical record exists. All records are subject to error tracing (logical mistakes such as

missing mandatory data, CPR-numbers of incorrect content, unknown hospital departments, etc.). In case of errors, the sender is requested to correct the record. Information from research projects is not routinely recorded in the registry.

A cornerstone of the DNPR and the DPDB is the use of the Danish version of Systemized Nomenclature of Medicine (SNOMED) codes. This system is based on the second edition of the College of American Pathologists 1979<sup>5</sup> which has been modified to incorporate Danish language terms, and updated regularly since. It follows the general principles of classification including the World Health Organization tumor classifications. Officially updated Danish SNOMED codes can be found at the DPDB Website (<http://www.patobank.dk/>) together with historical data, such as dates of code changes or modifications. Briefly, SNOMED codes are based on six axes, of which the two first are mandatory in all records: Topography (T), morphology (M), etiology (Æ), function (F), disease (S), and procedure (P). Each letter is followed by 5 digits determining the degree of detail in each axis. For more information on the use of SNOMED codes, see the DPDB Website. Table 2 contains a summary of the content of the DNPR according to overall T codes.

**Table 2** Overview of the number of specimens in the Danish National Registry of Patients categorized according to T-codes from the SNOMED system

Category of specimen	T-codes	1997–2000	2001–2004	2005–2008
Skin	T01–03	543820	692199	914303
Breast	T04	90567	107674	111239
Bone marrow	T06	87429	117575	140107
Spleen, lymph node, and thymus	T07–09/98	124697	178195	229621
Blood	T0x	43533	59533	58589
Bone, joint, muscle, tendons, and adipose tissue	T10–1x	35646	38708	39805
Respiratory tract	T21–29	176968	179444	184197
Heart and blood vessels	T31–49	30318	27143	26710
Mouth and salivary gland	T51–55	38226	41659	46964
Liver, bile ducts, and pancreas	T56–59/99	58548	65008	64773
Pharynx and tonsil	T60–61	11684	13984	15742
Esophagus and gastrointestinal tract	T62–69	358098	499036	605758
Kidney and urinary system	T71–75	94501	113444	128701
Male genitals	T76–79	206924	270203	405162
Female genitals	T80–87	471583	444071	450006
Placenta and foetus	T88–89	15705	17466	17611
Endocrine system, hypophysis, and pineal gland	T91–92	490	753	732
Adrenal gland, glomus, and paragangliomas	T93–96	1403	1753	2047
Thyroid and parathyroid gland	T96–97	24041	24940	29151
Nerve system (20)	Tx0–x9	21465	25791	26282
Eyes	Txx	2268	1805	1456
Ear	Txy	3573	3727	3534

**Notes:** The numbers are computed as the sum of all persons with one or more of the specific T-codes in each period. Hence, in this table, a patient's specimen counts only once per period in each category.

## Quality of the data

When evaluating the overall quality of the pathology data in the DNPR and the DPDB, it is important to consider whether all pathology examinations are recorded (no missing data) and whether the information recorded in the registry is correct. Information in the DNPR is obtained directly from the DPDB with theoretically minimal risk of data loss. The DPDB in turn functions as a central daily routine diagnostic tool for pathologists, with immediate direct entry of all results in the system and automatic online updating. Thus, the proportion of missing data in the DPDB (and in turn in the DNPR) is extremely low. However, before 1997 the degree of recorded details from pathology examinations varies (Table 1). It is difficult to assess the correctness of information in the DNPR and the DPDB (ie, the accuracy of the pathology conclusion and diagnosis for a particular investigation) because the pathology examination is considered the gold standard against which all clinical conclusions are compared. Moreover, the daily routines associated with pathology labs (including quality control and patient security) combined with the error tracing mentioned previously, reduce the risk of error and the effect of any such infrequent errors in research work should not be important.

## The use of pathology data in clinical research

Access to the DNPR or the DPDB provides detailed nationwide and population-based pathology information since 1997 and to some extent even before 1997 (Table 1).<sup>6,7</sup> By using the unique CPR number system, it is possible to link information from the DNPR or the DPDB to other research databases or population-based registries such as the National Registry of Patients. The National Registry of Patients has recorded all inpatient hospital diagnoses since 1977, and outpatient hospital diagnoses since 1995. Furthermore, data can also be used to validate other clinical databases or disease codes.<sup>8,9</sup>

## The use of biological specimens in clinical research

The data from the DNPR or the DPDB make it possible for researchers to identify histological and cytological specimens of interest from pathology archives throughout Denmark, from all patients since 1997 and to a more limited extent from before this date (Table 1). Since the records include information on the unique ID number of the specimens and the name of the pathology department holding the specimen, the precise physical location of the specimen can be

determined. Thus, it is possible to identify patients from throughout Denmark with a particular medical condition or disease of interest based on SNOMED codes, identify the location of the archived specimens, and retrieve these specimens for additional analyses. The results of the original and supplementary pathology analyses can be linked to data (including follow-up clinical data) available from other national and/or disease registers. The length of time for which pathology specimens are stored varies by pathology department and according to the type of biological specimen involved. Although not legally bound to do so, all Danish pathology departments store paraffin blocks permanently. In addition, some departments (usually those attached to the universities) have more limited diagnostic and/or research archives of frozen tissues that predate the establishment of the national frozen tissue biobank.

Access to archived pathology specimens is restricted, both for ethical reasons, and to preserve scarce specimens for future researchers. All research projects in Denmark that require access to biological specimens need to gain permission from the Danish Data Protection Agency and from the Danish National Committee on Biomedical Research Ethics. If the use of existing biobanks (including those at the pathology departments) does not involve any (new) risks or ethical considerations for the patients (which is usually the case in retrospective archive studies) dispensation for patient consent can be provided. Patients can specifically request that their specimens not be used in clinical research. This information is recorded in a national database, and this must be accessed by the researcher for each patient before their archived specimens can be used. Furthermore, the pathology departments involved have a duty to the patient to ensure that sufficient material is retained in the routine archive should the specimen be required later for further diagnostic analyses.

## How does one obtain access to the data and/or the biological specimens?

The DNPR and the DPDB are public registries and are, thus, available to researchers if the project and investigators in question meet the criteria determined by the Danish Data Protection Agency and the National Board of Health. To obtain data from the DNPR, an application should be sent to the National Board of Health through the online system on their Website (<http://www.sst.dk/>). To apply for access to the DPDB, one should consult their homepage (<http://www.patobank.dk/>). Both sites offer comprehensive application instructions in Danish. Simple data sampling can be



executed by the DPDB with short processing time while more complicated data sampling or data for the maintenance of clinical databases is carried out by the staff working with the DNPR in the Danish National Board of Health.

To access the biological specimens themselves, an additional approval from the Danish National Committee on Biomedical Research Ethics is necessary (<http://www.cvk.sum.dk/>). Furthermore, permission must be sought from each pathology department holding specimens that are to be included in a research project before its start. In general, Danish pathology departments are generous with their time and resources in cooperating with researchers even on projects in which they are not directly involved. However, it should be remembered that identifying, localizing, and extracting specimens from pathology archives is time-consuming work that may require specialist assistance from laboratory technicians and/or pathologists. This work will have a lower priority compared with the acute diagnostic duties of the departments, and may require some economic compensation.

## Conclusion

In the era of personalized medicine,<sup>10</sup> access to high quality pathology registries provides a valuable clinical and research resource. The DNPR and the DPDB facilitate research studies that are effectively immune to selection bias and representative of the target patient population.<sup>11</sup> Biomarkers of disease prognosis and biomarkers that predict treatment success may be investigated quickly and cost-effectively, and according to the highest quality standards.<sup>12</sup> In Denmark, access to well-documented archived (especially paraffin embedded) pathological specimens going back several decades makes it possible to perform large scale molecular epidemiological studies that can be efficiently linked to a wide range of other databases.

## Disclosure

The authors report no conflicts of interest in this work.

## References

1. Storm HH, Michelsen EV, Clemmensen IH, Pihl J. The Danish Cancer Registry-history, content, quality and use. *Dan Med Bull.* 1997;44(5):535-539.
2. Frank L. Epidemiology. When an entire country is a cohort. *Science.* 2000;287(5462):2398-2399.
3. Vyberg M, Bjerregaard B, Bak M, Gram I, Hvolris H. Pathology database. Danish Society of Pathologic Anatomy and Cytology [In Danish]. *Dan Med Bull.* 2005;167(12-13):1401.
4. Coordinating Group for Individual Patient Registration. Common content for registration of histopathology investigations. Report No.: 5, 3rd ed. 2002 [In Danish]. The Danish National Board of Health. Available from: [http://www.sst.dk/~media/Indberetning%20og%20statistik/Sundhedsstyrelsens%20registre/Patologiregister/Faellesindhold\\_patologi\\_2002.ashx](http://www.sst.dk/~media/Indberetning%20og%20statistik/Sundhedsstyrelsens%20registre/Patologiregister/Faellesindhold_patologi_2002.ashx). Accessed on January 11, 2010.
5. College of American Pathologists, Chicago Illinois. Systemized Nomenclature of Medicine (SNOMED). 2nd ed. 1979. Available from: <http://www.cap.org/>. Accessed on January 11, 2010.
6. Jepsen P, Vilstrup H, Sorensen HT. Alcoholic cirrhosis in Denmark – population-based incidence, prevalence, and hospitalization rates between 1988 and 2005: a descriptive cohort study. *BMC Gastroenterol.* 2008;8:3.
7. Friis S, Poulsen AH, Sorensen HT et al. Aspirin and other non-steroidal anti-inflammatory drugs and risk of colorectal cancer: a Danish cohort study. *Cancer Causes Control.* 2009;20(5):731-740.
8. Danish Clinical Database of Carcinomas of the Esophagus, Cardia and Stomach. Annual Report, 2004 [In Danish]. Available from: [http://www.kirurgisk-selskab.dk/dks/DECV/decv\\_aarsrapport/DECV1juni2003\\_31dec2004.pdf](http://www.kirurgisk-selskab.dk/dks/DECV/decv_aarsrapport/DECV1juni2003_31dec2004.pdf). Accessed on January 11, 2010.
9. Kjaerbye-Thygesen A, Huusom LD, Frederiksen K, Kjaer SK. Primary ovarian cancer. A comparison of registrations in the Danish Cancer Registry and the Pathology Data Bank [In Danish]. *Dan Med Bull.* 2007;169(1):50-54.
10. Donovan MJ, Costa J, Cordon-Cardo C. Systems pathology: a paradigm shift in the practice of diagnostic and predictive pathology. *Cancer.* 2009;115(13 Suppl):3078-3084.
11. Ransohoff DF, Gourlay ML. Sources of bias in specimens for research about molecular markers for cancer. *J Clin Oncol.* 2010;28(4):698-704.
12. Simon RM, Paik S, Hayes DF. Use of archived specimens in evaluation of prognostic and predictive biomarkers. *J Natl Cancer Inst.* 2009;101(21):1446-1452.

### Clinical Epidemiology

### Publish your work in this journal

Clinical Epidemiology is an international, peer-reviewed, open access 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

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

Dovepress

reviews, risk & safety of medical interventions, epidemiology & biostatistical methods, evaluation of guidelines, translational medicine, health policies & 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.