



## Validation of The Danish Cancer Registry and selected Clinical Cancer Databases

### *Abstract*

**Background/introduction:** The nationwide Danish Cancer Registry (DCR) underwent a comprehensive modernization from 2004 to 2008. From 1943 to 2003 all coding was based on paper notifications and performed manually by specially trained personnel. Since 2004, the modernized DCR has been automatized through linkage between the National Patient Register, the Danish Pathology Registry and the Danish Register of Causes and Death, with the Patient Register as primary data source. In addition, a separate electronic reporting system has been developed for cancer reporting from the primary health care sector, including private clinics. Reports of cancer diagnoses in the Patient Register is linked to information from the Pathology Registry for addition of pathological details and validation, using an automatic coding system which converts relevant data for the patient to individual-based tumor records. The key information in the modernized DCR is similar to the information in the former notification-based registry to secure historical perspectives. For completeness, the modernized DCR depends on manual processing of a small proportion of cancer cases and for validation purposes.

To evaluate the quality of the modernized DCR, The National Board of Health, the Danish Cancer Society and representatives from the Danish Lung Cancer Group (DLCG) and Danish Breast Cancer Cooperative Group (DBCG) performed a validation study, with the primary focus of assessing the precision and completeness of the DCR. A secondary purpose was to evaluate the performance of the clinical databases.

**Methods:** From the DCR, information on all recorded cases of lung and breast cancer during 2005-2007 was retrieved. The index year of validation was 2006, whereas 2005 and 2007 data were used to accommodate and validate cancer diagnoses close to turn of the year. The information included individual-level data on personal identification number, diagnosis(es) according to the International Classification of Diseases, version 10 (ICD-10), date of diagnosis, and pathological details, including topography (specific localization) and morphology (histological type) according to the ICD for oncology (ICD-O). The information from the DCR was compared to the relevant data on lung and breast cancer from the DLCR and DBCG clinical databases. In case of disparity between the Cancer Registry and clinical databases or missing records in either the DCR or the clinical databases, complete records were retrieved from the Pathology Registry and the Patients Register for diagnostic clarification. For cases that remained unresolved after the additional evaluation, medical records were retrieved from the relevant hospital departments and evaluated manually by the study group. To estimate the number of incident cancer cases missing in either the Cancer Registry or the relevant database a 'capture-recapture analysis' was performed.

**Results:** Among 4443 lung cancer cases registered in either the DCR or the DLCR in 2006, 732 cases exhibited important inconsistencies between the two data sources. Thirteen (12.9) percent of valid cases of lung cancer were either missing or recorded with an incorrect diagnosis in DLCR, while the corresponding proportion was 3.6 % in the DCR. It was estimated that 14 cases of lung cancer were not captured in either registry.



For invasive breast cancer, 374 cases of important inconsistencies were found among 4411 cases in the DCR and/or the DBCG. The proportion of cases missing or with incorrect diagnosis was 7.1% in the DLCR and 1.2 % in the DCR. It was estimated that four cases of invasive breast cancer were not captured by either data sources. For in situ breast cancer, the corresponding figures were 9.5 % and 1.8%, with one not captured case.

Conclusion: The results of the validation study were generally satisfactory, revealing low proportions of errors or missing reports of lung and breast cancer in the DCR. The higher precision in the recording of breast cancer cases than of lung cancer was probably due to differences in the primary diagnostics. The new automatized DCR ensures continued high quality and completeness in registration of cancer in Denmark. The DCR is a dynamic registry and continued improvements in the coding are being performed on a continuous basis and corrected back in time. Minor improvements in the cancer registration were also achieved by the present validation study.