

# Existing data sources for clinical epidemiology: The clinical laboratory information system (LABKA) research database at Aarhus University, Denmark

Anne Fia Grann  
Rune Erichsen  
Anders Gunnar Nielsen  
Trine Frøslev  
Reimar W Thomsen

Department of Clinical Epidemiology,  
Aarhus University Hospital,  
Aarhus, Denmark

**Abstract:** This paper provides an introduction to the clinical laboratory information system (LABKA) research database in Northern and Central Denmark. The database contains millions of stored laboratory test results for patients living in the two Danish regions, encompassing 1.8 million residents, or one-third of the country's population. More than 1700 different types of blood test analyses are available. Therefore, the LABKA research database represents an incredible source for studies involving blood test analyses. By record linkage of different Danish registries with the LABKA research database, it is possible to examine a large number of biomarkers as predictors of disease risk and prognosis and as markers of disease severity, and to evaluate medical treatments regarding effectiveness and possible side effects. Large epidemiological studies using routinely stored blood test results for individual patients can be performed because it is possible to link the laboratory data to high-quality individual clinical patient data in Denmark.

**Keywords:** biochemistry, laboratory procedures, diagnosis, therapeutic drug monitoring, epidemiological methods, registries

## Introduction

Every day, millions of blood specimens are analyzed as part of routine clinical work in hospitals and medical practices throughout the world. Laboratory tests performed are analyzed in clinical laboratories and are usually recorded in computer-based laboratory information systems.<sup>1</sup>

The clinical laboratory information system (LABKA) research database in Denmark contains millions of stored laboratory test results from Danish patients living in the North Denmark region and Central Denmark region, representing one-third of the country's population. Therefore, the LABKA research database represents an incredible source for studies involving blood test analyses. Large epidemiological studies can be performed, if the laboratory data can be linked to high-quality clinical patient data at the individual level, as is 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 recording periods, high quality, and completeness.<sup>2,3</sup> Widespread 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 tax-funded medical care for all 5.5 million Danish residents. Since 1968, all Danish residents have been assigned a unique personal identifier, ie, the civil personal registration (CPR) number, which is used in all health and population databases and

Correspondence: Anne Fia Grann  
Department of Clinical Epidemiology,  
Institute of Clinical Medicine, Aarhus  
University Hospital, Olof Palmes Allé  
43-45, DK-8200 Aarhus N, Denmark  
Tel +45 8942 4800  
Fax +45 8942 4801  
Email afg@dce.au.dk

facilitates unambiguous computerized record linkage. In this paper, we present an overview of the clinical laboratory information system and the LABKA research database. Our objective is that researchers understand the Danish laboratory information system and learn some of the potential uses that can be made of the Danish biochemistry data in clinical research.

## Danish clinical laboratory information systems

Today, computer-based laboratory information systems are functioning as a central routine diagnostic tool for medical doctors in all private clinics and hospital departments in Denmark. New laboratory test results are entered immediately and directly into the system, with automatic online updating and access to results for all relevant hospital personnel. In 1979, the first version of a real-time computer-based laboratory information system, ie, LABKA, was designed and implemented at Rigshospitalet in Copenhagen, initially for internal use in the clinical chemistry department only.<sup>4</sup> Since then, the development of the LABKA system has continued.

In 1985, LABKA was implemented as part of the daily routine in all clinical departments at the Aarhus Amtssygehus (Central Denmark region) as the first commercial laboratory information system. Gradually, LABKA expanded and today the most recent system, called LABKA II,<sup>5</sup> is implemented in all clinical chemistry departments in public hospitals (more than 40) in three of the five Danish administrative regions, ie, the North Denmark region, the Central Denmark region, and the Copenhagen region (data from the latter region is not included in the LABKA research database, see below). The LABKA system was developed by the Dansk Datalab, later renamed the CSC-Datalab.<sup>5</sup>

The LABKA system holds test results from every blood sample taken in any public or private hospital or by any general practitioner and submitted to any clinical chemistry department located in the three Danish regions. Exceptions are some results from small and rapid point-of-care devices used by medical staff or patients themselves for instant analysis of, eg, INR, blood glucose, hemoglobin, and C-reactive protein (CRP).<sup>6,7</sup> When a blood sample has been analyzed in the clinical chemistry department laboratory, the sample is destroyed and the results are electronically transmitted to the requesting hospital department or general practitioner through the LABKA system. The information is recorded in a uniform way according to the international NPU (Nomenclature, Properties and Units) coding system.

The NPU code is the unique identification number for each single investigation. The NPU coding system provides a terminology for identification of clinical laboratory test values following the international recommendations for efficient electronic communication in clinical laboratories, and ensures that names and units are shown in a uniform manner.<sup>8</sup>

## LABKA research database

The laboratory information systems provide data for the LABKA research database. Traditionally, it has been called “the LABKA database” after the commercial name of the laboratory information system. A more correct name might have been the “laboratory information system database”. The population covered by the database comes from a well-defined geographical area corresponding to two of the five Danish administrative regions, ie, the North Denmark region and the Central Denmark region, encompassing 1.8 million residents. The regions consist of mixed rural and urban areas with modestly growing populations.

Hospitals started transferring data to the LABKA database in different time periods, beginning in 1990. The database is considered to have full geographical coverage in the North Denmark region since 1997 and in the Central Denmark region since 2000. The LABKA research database is currently updated once a year, and is managed by the Department of Clinical Epidemiology at Aarhus University Hospital.

## Information recorded in the LABKA database

As previously indicated, all results of analyzed blood samples drawn from patients from hospital departments and medical practices in the North and Central Denmark regions are recorded in the LABKA research database according to NPU codes and specific analysis numbers. In addition, date of blood sample analysis, patient CPR number, and identification code for the physician or hospital department requesting the blood test analysis are recorded.

Table 1 shows the number of tests recorded in the LABKA research database for some of the most common analyses used in everyday clinical practice.

Around 1700 different blood sample analyses are available in the database. Among these are 76 pharmacological analyses. Table 2 shows selected examples from the LABKA database of a number of pharmacological analyses where monitoring is recommended as a guide to treatment. Analyzed blood samples are by far the most common type

**Table 1** NPU codes and numbers of analyzed blood tests from January 1, 1999 to December 31, 2007, showing some of the most frequent blood test analyses in the LABKA database

Blood test	NPU code	1999–2001	2002–2004	2005–2007	Total blood tests	Patients in total*
Sodium	03429	2,139,244	2,708,127	3,352,908	8,200,279	938,173
Potassium	03230	2,202,088	2,838,745	3,462,178	8,503,011	950,680
Creatinine	01807	2,273,490	2,867,271	3,104,634	8,245,395	899,050
	18016					
	18105					
Urea	01459	1,103,001	1,190,903	1,183,272	3,477,176	490,401
Albumin	01132	1,309,050	1,674,173	2,237,524	5,220,747	771,107
	19673					
	02187					
Glucose	02187	1,276,282	1,470,945	1,707,899	4,455,126	590,015
	02192					
	02193					
	02195					
	04093					
	08520					
	08567					
	08869					
	08916					
	08972					
	22069-					
	22089					
	22095					
	22099-					
	22126					
Glycosylated hemoglobin	02307	277,898	415,222	582,800	1,275,920	298,777
	03835					
	27300					
Hemoglobin	02319	2,209,214	2,735,613	3,288,734	8,233,561	1,011,861
	02322					
Leukocytes	02593	2,324,157	3,065,182	3,217,612	8,606,951	951,268
	02596					
	04100					
	18156					
C-reactive protein	01423	1,277,912	1,686,060	1,775,578	4,739,550	781,110
	05027					
	19748					
Total	–	16,392,336	20,652,241	23,913,139	60,957,716	

**Notes:** \*The numbers are computed as the sum of all blood tests in each period. In this table, a person can have one or more blood test results in each period.

**Abbreviation:** NPU, Nomenclature, Properties and Units.

of analysis in the LABKA research database. However, other specimen analysis types such as urine and fecal occult blood tests, and cerebrospinal fluid are also accessible in the database.

## Data quality

When evaluating the overall quality of the laboratory data in the LABKA research database, it is important to consider whether all laboratory data are recorded and whether the information recorded in the registry is correct. Information in the LABKA research database is obtained directly, electronically, and without filtering from the LABKA systems in the North Denmark region and the Central Denmark region, theoretically with minimal risk of data loss. Because the LABKA

systems in the two regions function as an indispensable daily routine diagnostic tool for all medical personnel, and is based on immediate direct entry of approved results into the system, the proportion of missing data in the LABKA system (and in turn in the LABKA research database) is extremely low. Errors due to technical faults are hard to detect, but are the same errors that may be seen in everyday clinical practice. Unsuccessful and cancelled blood samples are also recorded in the database, and therefore can be excluded before using LABKA data in research. Furthermore, in the different hospital laboratories, different medical equipment may have been used over time, and variation in, eg, units and normal ranges may therefore occur for some analyses. Finally, different local analysis codifications have been used for a

**Table 2** Examples of a number of pharmacological analyses where monitoring is indicated as a guide to treatment

Blood test	NPU code	1999–2001	2002–2004	2005–2007	Total blood tests	Patients in total*
Clozapine	04114	3660	5535	20,004	29,199	10,437
Lamotrigine	08732	8607	13,075	17,143	38,825	8075
Valproate	03735	12,147	12,904	13,577	38,628	7018
	14378					
Nortriptyline	02923	3385	4954	4925	13,264	3817
Lithium	02613	20,318	20,867	20,420	61,605	4189
Paracetamol	03024	4005	6002	5584	15,591	9290
	16789					
	18232					
Salicylate	03383	3290	4339	4258	11,887	6587
Gentamicin	02164	2164	6609	8645	17,418	6581
	19779					
Vancomycin	19929	203	1090	2552	3845	529
Methotrexate	02739	3244	3287	3556	10,087	736
Cyclosporine	01592	17,116	14,938	16,532	48,586	2995
	03990					
	19725					
	19726					
Digoxin	01886	43,914	32,255	29,061	105,230	26,920
INR	01685	720,035	922,822	983,899	2,626,756	300,527
Total	–	842,088	1,048,677	1,130,156	3,020,921	

**Notes:** \*The numbers are computed as the sum of all blood tests in each period. In this table, a person can have one or more blood test results in each period.

number of tests in the early years of the database before the uniform NPU coding was implemented in the laboratories in 2001. Many of these previous analysis codes have been transformed into NPU codes, but this transformation is not fully complete for all laboratory tests. However, correct transformation of different codes and units as needed by the researcher is possible with the information available in the LABKA research database.

## Data linkage possibilities

All patients are registered by their individual CPR number, and this is one of the database's major strengths. Use of the CPR number, assigned to all Danes, makes it possible to link laboratory data to a wide range of other medical and socioeconomic Danish registries. This includes the Danish national registry of patients (with data on all hospital admissions and discharge diagnoses since 1977 and all outpatient visits since 1995),<sup>9</sup> the Danish civil registration system (which contains exact dates of immigration, emigration, and death, if any), and the prescription database at Aarhus University.<sup>10</sup> Hereby, it is possible to build large cohorts with detailed longitudinal data that include full hospitalization history and outpatient visits, comorbidity data, prescriptions redeemed at any pharmacy, and detailed cumulated laboratory results. The population base makes the database well suited for matched cohort and case-control studies, with the Danish civil registration system facilitating the selection of

controls from the underlying population using incidence density sampling.<sup>11</sup>

Examples of studies involving linkage between the LABKA database and other data sources are given below. Even unlinked data from the LABKA database can provide a wealth of information, but linking data with, eg, the Danish national registry of patients, may provide researchers with invaluable data on the course of acute as well as chronic diseases.

For a number of drugs, monitoring of therapeutic ranges is recommended as a guide to treatment, eg, antipsychotic, antibiotic, antiepileptic, or immunosuppressive treatment securing optimal drug effect, minimal side effects, compliance, and dosage of medical treatments. Pharmacological analyses are also used when diagnosing drug poisoning.

The prescription database at Aarhus University collects data on reimbursed medications dispensed by all community pharmacies in the North Denmark region and Central Denmark region, ie, the same regions as in the LABKA database. Combining pharmacy records with laboratory data can generate important studies in clinical epidemiology, including drug utilization studies, safety monitoring, and etiologic research. LABKA data can provide up-to-date information on treatment status, and thereby evaluate medical therapies in both hospitals and in primary care. It is desirable to develop methods that reflect current clinical practice in large geographical areas and to be able to study

the treatment status for different kinds of drugs, by use of information from the LABKA research database and other existing registries.

## Examples of studies using LABKA data

Biomarkers from the LABKA database may be investigated as prognostic predictors. In a recent cohort study, we linked patients with bacteremia identified in the North Denmark bacteremia research database for the 1997–2004 period<sup>12</sup> with the LABKA database to examine the association between the CRP level at the time of blood culture draw and subsequent bacteremia mortality. We were able to identify CRP measurements, including exact date, hour, and minute of drawing the CRP specimen, for 5449 (95.1%) of 5730 bacteremia patients.<sup>13</sup> Thirty-day mortality increased consistently from 13.2% in the first CRP quartile (10–64 mg/L) to 24.8% in the fourth CRP quartile (241–688 mg/L). After controlling for confounding factors, the baseline CRP level remained an independent predictor for 30-day mortality. The adjusted 30-day mortality rate ratios for patients in the second, third, and fourth CRP quartiles were 1.38 (95% confidence interval [CI] 1.13–1.69), 1.70 (95% CI 1.40–2.06), and 2.38 (95% CI 1.96–2.87), respectively.<sup>13</sup>

Biomarkers from the LABKA database may also be investigated as treatment outcomes. A study from 2007 examined the association between aminoglycoside therapy and increase in serum creatinine among patients with bacteremia.<sup>14</sup> Serum creatinine was available at baseline for 77.9% of patients treated with aminoglycosides and 79.2% of patients not treated with aminoglycosides. An increase in serum creatinine of  $\geq 45$  mmol/L within 14 days after the bacteremia was observed almost as often in aminoglycoside-treated and non-aminoglycoside-treated patients (14.1% versus 12.4%, adjusted odds ratio 1.06; 95% CI 0.63–1.79). We could prove that this finding was partly due to aminoglycosides being less frequently used in patients with high baseline creatinine levels (who had the highest risk of further increases in creatinine).<sup>14</sup>

Biomarkers from the LABKA database may further be investigated as markers of underlying disease. Kornum et al conducted a large case-control study to examine whether diabetes and glycemic control assessed by HbA1c level is a risk factor for hospitalization with pneumonia.<sup>15</sup> Their 34,239 pneumonia patients and 342,390 population controls were linked to the LABKA database. One or more HbA1c measurements within 12 months were available for 2731

(61%) of pneumonia patients with diabetes and for 16,605 (58%) of control subjects with diabetes. The most recent HbA1c measurement was used in the analysis. Compared with subjects without diabetes, the adjusted relative risk for pneumonia was 1.22 (95% CI 1.14–1.30) for diabetic subjects whose HbA1c level was  $<7\%$  and 1.60 (95% CI 1.44–1.76) for diabetic subjects whose HbA1c level was  $\geq 9\%$ . The authors earlier demonstrated that inflammatory markers (including CRP and leukocytes) on admission with pneumonia were similarly high in diabetic and other individuals, suggesting that there was no major hospital referral bias associated with having diabetes.<sup>16</sup>

## Accessing the LABKA data

Data files are kept for research purposes by the Department of Clinical Epidemiology, while the original data files are kept by the North Denmark region and Central Denmark region. The Danish Act on Processing of Personal Data (Persondataloven) provides the legal basis for the ability of public institutions, including universities, to retain person-identifiable health data for research purposes. In order to access the data from the LABKA research database, researchers are welcome to contact Hanne K Schlosser hks@dce.au.dk at the Department of Clinical Epidemiology, Aarhus University Hospital (www.kea.au.dk). There is a steering committee for the database. Furthermore, use of any health data requires project-specific permission from the Data Protection Agency (Datatilsynet, www.datatilsynet.dk).

## Conclusion

The LABKA research database in Denmark contains many millions of stored laboratory test results covering 1.8 million people for a period of more than 10 years. These data may facilitate research within all areas of medicine. By record linkage with other Danish databases and registries, it is possible to conduct large clinical epidemiologic studies to examine important biomarkers as predictors of disease risk and prognosis, and to evaluate medical treatments regarding efficacy and possible side effects.

## Acknowledgments

We would like to thank Lars Pedersen, Lars Ulrik Gerdes, and Niels Jørgen Christensen for their helpful suggestions and contributions to this research.

## Disclosure

The authors report no conflicts of interest for this work.

## References

1. McQuilten ZK, Schembri N, Polizzotto MN, et al. Hospital blood bank information systems accurately reflect patient transfusion: Results of a validation study. *Transfusion (Paris)*. November 15, 2010. [Epub ahead of print].
2. Frank L. Epidemiology. When an entire country is a cohort. *Science*. 2000;287:2398–2399.
3. Storm HH, Michelsen EV, Clemmensen IH, et al. The Danish Cancer Registry – history, content, quality and use. *Dan Med Bull*. 1997;44: 535–539.
4. Christiansen JU, Maruard CD, Nielsen HC. LABKA. A real-time computer system for the clinical laboratory. *Scand J Clin Lab Invest Suppl*. 1989;194:57–61.
5. CSC. Laboratoriesystemer. Available at: <http://www.scandihealth.dk/Losninger/Laboratoriesystemer.htm>. Accessed March 10, 2011.
6. Hoedemaekers CW, Klein Gunnewiek JM, Prinsen MA, et al. Accuracy of bedside glucose measurement from three glucometers in critically ill patients. *Crit Care Med*. 2008;36:3062–3066.
7. Kitchen DP, Kitchen S, Jennings I, et al. Quality assessment in point-of-care coagulation testing. *Semin Thromb Hemost*. 2008;34:647–653.
8. Pontet F, Magdal PU, Fuentes-Arderiu X, et al. Clinical laboratory sciences data transmission: The NPU coding system. *Stud Health Technol Inform*. 2009;150:265–269.
9. Andersen TF, Madsen M, Jorgensen J, et al. The Danish National Hospital Register. A valuable source of data for modern health sciences. *Dan Med Bull*. 1999;46:263–268.
10. Gaist D, Sorensen HT, Hallas J. The Danish prescription registries. *Dan Med Bull*. 1997;44:445–448.
11. Wacholder S, Silverman DT, McLaughlin JK, et al. Selection of controls in case-control studies. III. Design options. *Am J Epidemiol*. 1992;135: 1042–1050.
12. Schonheyder HC, Sogaard M. Existing data sources for clinical epidemiology: The North Denmark Bacteremia Research Database. *Clin Epidemiol*. 2010;2:171–178.
13. Gradel KO, Thomsen RW, Lundbye-Christensen S, et al. Baseline C-reactive protein level as a predictor of mortality in bacteraemia patients: A population-based cohort study. *Clin Microbiol Infect*. 2010.
14. Freundlich M, Thomsen RW, Pedersen L, et al. Aminoglycoside treatment and mortality after bacteraemia in patients given appropriate empirical therapy: A Danish hospital-based cohort study. *J Antimicrob Chemother*. 2007;60:1115–1123.
15. Kornum JB, Thomsen RW, Riis A, et al. Diabetes, glycemic control, and risk of hospitalization with pneumonia: A population-based case-control study. *Diabetes Care*. 2008;31:1541–1545.
16. Kornum JB, Thomsen RW, Riis A, et al. Type 2 diabetes and pneumonia outcomes: A population-based cohort study. *Diabetes Care*. 2007;30: 2251–2257.

### 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.