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.1

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.2,3 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...
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. 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, 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.5

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).6,7 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.5

**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...
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 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

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Sodium</td>
<td>03429</td>
<td>2,139,244</td>
<td>2,708,127</td>
<td>3,352,908</td>
<td>8,200,279</td>
<td>938,173</td>
</tr>
<tr>
<td>Potassium</td>
<td>03230</td>
<td>2,202,088</td>
<td>2,838,745</td>
<td>3,462,178</td>
<td>8,503,011</td>
<td>950,680</td>
</tr>
<tr>
<td>Creatinine</td>
<td>01807</td>
<td>2,273,490</td>
<td>2,867,271</td>
<td>3,104,634</td>
<td>8,245,395</td>
<td>899,050</td>
</tr>
<tr>
<td>Urea</td>
<td>01459</td>
<td>1,103,001</td>
<td>1,190,903</td>
<td>1,183,272</td>
<td>3,477,176</td>
<td>490,401</td>
</tr>
<tr>
<td>Albumin</td>
<td>01132</td>
<td>1,309,050</td>
<td>1,674,173</td>
<td>2,237,524</td>
<td>5,220,747</td>
<td>771,107</td>
</tr>
<tr>
<td>Glucose</td>
<td>02187</td>
<td>1,276,282</td>
<td>1,470,945</td>
<td>1,707,899</td>
<td>4,455,126</td>
<td>590,015</td>
</tr>
<tr>
<td>Glycosylated hemoglobin</td>
<td>02307</td>
<td>277,898</td>
<td>415,222</td>
<td>582,800</td>
<td>1,275,920</td>
<td>298,777</td>
</tr>
<tr>
<td>Hemoglobin</td>
<td>02319</td>
<td>2,209,214</td>
<td>2,735,613</td>
<td>3,288,734</td>
<td>8,233,561</td>
<td>1,011,861</td>
</tr>
<tr>
<td>Leukocytes</td>
<td>02593</td>
<td>2,324,157</td>
<td>3,065,182</td>
<td>3,217,612</td>
<td>8,606,951</td>
<td>951,268</td>
</tr>
<tr>
<td>C-reactive protein</td>
<td>01423</td>
<td>1,277,912</td>
<td>1,686,060</td>
<td>1,775,578</td>
<td>4,739,550</td>
<td>781,110</td>
</tr>
<tr>
<td>Total</td>
<td>–</td>
<td>16,392,336</td>
<td>20,652,241</td>
<td>23,913,139</td>
<td>60,957,716</td>
<td>1,780,862</td>
</tr>
</tbody>
</table>

**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.
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),\(^9\) the Danish civil registration system (which contains exact dates of immigration, emigration, and death, if any), and the prescription database at Aarhus University.\(^10\) 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.\(^11\)

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

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Clozapine</td>
<td>04114</td>
<td>3660</td>
<td>5535</td>
<td>20,004</td>
<td>29,199</td>
<td>10,437</td>
</tr>
<tr>
<td>Lamotrigine</td>
<td>08732</td>
<td>8607</td>
<td>13,075</td>
<td>17,143</td>
<td>38,825</td>
<td>8075</td>
</tr>
<tr>
<td>Valproate</td>
<td>03735</td>
<td>12,147</td>
<td>12,904</td>
<td>13,577</td>
<td>38,628</td>
<td>7018</td>
</tr>
<tr>
<td>Noroxetine</td>
<td>02923</td>
<td>3385</td>
<td>4954</td>
<td>4925</td>
<td>13,264</td>
<td>3817</td>
</tr>
<tr>
<td>Lithium</td>
<td>02613</td>
<td>20,318</td>
<td>20,867</td>
<td>20,420</td>
<td>61,605</td>
<td>4189</td>
</tr>
<tr>
<td>Paracetamol</td>
<td>03024</td>
<td>4005</td>
<td>6002</td>
<td>5584</td>
<td>15,591</td>
<td>9290</td>
</tr>
<tr>
<td>Salicylate</td>
<td>03383</td>
<td>3290</td>
<td>4339</td>
<td>4258</td>
<td>11,887</td>
<td>6587</td>
</tr>
<tr>
<td>Gentamicin</td>
<td>02164</td>
<td>2164</td>
<td>6609</td>
<td>8645</td>
<td>17,418</td>
<td>6581</td>
</tr>
<tr>
<td>Vancomycin</td>
<td>19929</td>
<td>203</td>
<td>1090</td>
<td>2552</td>
<td>3845</td>
<td>529</td>
</tr>
<tr>
<td>Methotrexate</td>
<td>02739</td>
<td>3244</td>
<td>3287</td>
<td>3556</td>
<td>10,087</td>
<td>736</td>
</tr>
<tr>
<td>Cyclosporine</td>
<td>01592</td>
<td>17,116</td>
<td>14,938</td>
<td>16,332</td>
<td>48,586</td>
<td>2995</td>
</tr>
<tr>
<td>Digoxin</td>
<td>01886</td>
<td>43,914</td>
<td>32,255</td>
<td>29,061</td>
<td>105,230</td>
<td>26,920</td>
</tr>
<tr>
<td>INR</td>
<td>01685</td>
<td>720,035</td>
<td>922,822</td>
<td>983,899</td>
<td>2,626,756</td>
<td>300,527</td>
</tr>
<tr>
<td>Total</td>
<td>–</td>
<td>842,088</td>
<td>1,048,677</td>
<td>1,130,156</td>
<td>3,020,921</td>
<td></td>
</tr>
</tbody>
</table>

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.

Table 2 Examples of a number of pharmacological analyses where monitoring is indicated as a guide to treatment
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 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. Thirty-day mortality, hour, and minute of drawing the CRP specimen, for 5449 (95.1%) of 5730 bacteremia patients. 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.

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. 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 ≥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).

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. 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 ≥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.

**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


