

[Home](#) > [Science Magazine](#) > [31 March 2000](#) > Frank, 287 (5462): 2398-2399

Science 31 March 2000:  
Vol. 287 no. 5462 pp. 2398-2399  
DOI: 10.1126/science.287.5462.2398

• NEWS FOCUS

EPIDEMIOLOGY

**When an Entire Country Is a Cohort**

[Prev](#) | [Table of Contents](#) | [Next](#)

Lone Frank\*

Lone Frank writes from Copenhagen, Denmark.

**Denmark has gathered more data on its citizens than any other country. Now scientists are pushing to make this vast array of statistics even more useful**

For years, any woman who got an abortion had to accept more than the loss of her fetus: For some unknown reason, she also faced an elevated risk for breast cancer. At least that was what several small case-control studies had suggested before Mads Melbye, an epidemiologist at the Statens Serum Institute in Copenhagen, undertook

the largest effort ever to explore the link. He and his colleagues obtained records on 400,000 women in Denmark's national Abortion Register, then checked how many of the same women were listed in the Danish Cancer Register. Their foray into the two databases led to a surprising result: As they reported in *The New England Journal of Medicine* in 1997, there appears to be no connection between abortion and breast cancer.

Their success underscores the value of a trove of data the Danish government has accumulated on its citizenry, which today totals about 5 million people. Other Scandinavian countries have created powerful database systems, but Denmark has earned a preeminent reputation for possessing the most complete and interwoven collection of statistics touching on almost every aspect of life. The Danish government has compiled nearly 200 databases, some begun in the 1930s, on everything from medical records to socioeconomic data on jobs and salaries. What makes the databases a plum research tool is the fact that they can all be linked by a 10-digit personal identification number, called the CPR, that follows each Dane from cradle to grave. According to Melbye, "our registers allow for instant, large cohort studies that are impossible in most countries."

But Melbye and other scientists think they can extract even more from this data gold mine. They argue that not enough money is being spent on maintaining and expanding existing databases, and they say that red tape is hampering studies that require correlation of health and demographic data. The problem is that, while they have unfettered access to more than 80 medical databases maintained

by the Danish Board of Health and public hospitals, their use of 120 demographic databases overseen by the agency Statistics Denmark is tightly restricted. Statistics Denmark won't allow researchers to remove from its premises data coded by CPR, and the procedures for accessing information at all are unwieldy and expensive.

Statistics Denmark officials are reluctant to release data tied to CPRs, citing privacy concerns. "The public should have confidence that information identifying them as individuals does not reside outside of this institution," says the agency's Otto Andersen. Last month, Danish research minister Birte Weiss formed a committee to break the impasse. Denmark's databases are "a resource which can be used more optimally," she told *Science*. "This should be a scientific flagship."

Working the health databases can yield powerful results. For years the U.S. National Institutes of Health has supported a study following twins, hoping to tease out the relative contributions of genes and lifestyle to aging. Led by University of Southern Denmark gerontologist Kaare Christensen, the project has tapped the Danish Twin Register, which includes 110,000 pairs of twins born since 1870. After following more than 2000 pairs of twins aged 70 or older, Christensen's group has so far tied to genes about a quarter of the variation in human longevity. "The project is made possible by the unmatched age and completeness of the Danish Twin Register," he says.

The health databases have proven invaluable for probing contradictions raised by smaller studies and following disease progression. Tapping the Danish Cancer Register and a central blood bank database, Melbye's team recently debunked the idea that getting a blood transfusion increases cancer risk. And in another large project, they resolved the long-standing question of why young women with breast cancer have a poorer prognosis than older women with the disease. Compiling data on 35,000 breast cancer cases, the researchers found that young victims live longer if chemotherapy is started as soon as the cancer is diagnosed, rather than reserving this treatment approach only for tumors that have grown to a certain size.

The health databases are also useful for unraveling complex diseases. Psychiatric epidemiologist Preben Bo Mortensen of Aarhus University Hospital has mined the Danish Central Psychiatric Register, which contains information on all Danes who have come into contact with the public psychiatric hospital system since the 1930s. He has identified a host of environmental factors, such as prenatal viral infections and season of birth, which appear to influence the development of schizophrenia and bipolar disorder. "The register allows us to tease out the relative contribution of genetic and nongenetic factors and thereby point to possible strategies for preventing disease," says Mortensen.

But at a meeting on database research in Copenhagen earlier this year, scientists complained that they have been prevented from taking full advantage of the wealth of registered information. Part of the problem is that the agencies that maintain databases are reeling from budget cuts. "Increased funds are needed to secure the quality of existing and future databases if we want to keep our lead in the field," says Olaf Ingerslev of the Board of Health. Experts estimate it would take only a modest additional cash infusion—\$500,000 to \$1 million a year—to better maintain and expand the Board of Health's databases.

Budget shortfalls, however, are not as contentious as the question of access. "The central issue is the refusal by Statistics Denmark to release personally identifiable CPR-related data," says epidemiologist Thorkild I. A. Sørensen of Copenhagen's Institute for Disease Prevention. The agency's rules, stricter than mandated by law, make accessing data a cumbersome process. If researchers want to link data from a health database to data from a demographic database, for example, they pay a steep fee for an appointment at Statistic Denmark's Copenhagen office, where they wait while a bureaucrat carries out the request. Researchers can't take home with them data coded by CPR, which constrains how they manipulate data at their institutions. Performing follow-up analyses that require linking data by CPRs means returning to Statistics Denmark and paying another fee.

Researchers argue that the benefits of entrusting them with the CPR outweigh the risk of compromising the identifying information. It's "better to meet concerns by tightening possible sanctions than to limit research that benefits society in general," says Sørensen. While the company deCODE's plan to create and mine an Icelandic

health database has provoked heated debate in Iceland and abroad (*Science*, 11 February, p. [951](#)), the database issue has aroused little concern in Denmark. In part that's because no one has suffered the embarrassment of having their medical records inadvertently released into the public domain: Researchers must strip CPRs from health data before publishing analyses, and they have no access to the names that go with the CPRs.

As the government committee considers their request, the researchers are forging ahead with new projects that aim to marry advances in genetics with the vast database resources. "The ability to track related individuals in the many different databases makes it possible to shed light on the complex interplay between familial predisposition and environment," says Melbye. Christensen's twins, for example, donate blood samples that are used to analyze genes implicated in aging. And in the Danish National Birth Cohort, 100,000 pregnant women and their babies are being recruited to donate blood and undergo physical exams during pregnancy. The emerging database may point to new connections between prenatal factors and congenital disorders, as well as chronic diseases that occur later in life. "In the future we can go back and analyze data concerning the mothers' health during pregnancy and test suspected genetic and nongenetic factors in blood," says epidemiologist Jørn Olsen of Aarhus University, one of the project's leaders. The data could be all the more useful, he says, if the current restraints on linking information across databases are lifted. Indeed, Christensen sees it as a moral obligation to exploit data gathered at great expense. "Would it not be unethical not to use it to improve the population's health and health care?" he asks.

---

**THIS ARTICLE HAS BEEN CITED BY OTHER ARTICLES:**

**Thromboembolic risk among Danish children and adults with inflammatory bowel diseases: a population-based nationwide study**

*Gut* 1 July 2011: 937-943.

[Abstract](#) [Full Text](#) [Full Text \(PDF\)](#)

**The Danish National Prescription Registry**

*Scand J Public Health* 1 July 2011: 38-41.

[Abstract](#) [Full Text \(PDF\)](#)

**A Prospective Cohort Study of Menstrual Characteristics and Time to Pregnancy**

*Am J Epidemiol* 30 June 2011: kwr130v1-kwr130.

[Abstract](#) [Full Text](#) [Full Text \(PDF\)](#)