Cohort Profile Update: The Danish HIV Cohort Study (DHCS)

Lars Haukali Omland,* Magnus Glindvad Ahlström and Niels Obel

Department of Infectious Diseases, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark

*Corresponding author. Department of Infectious Diseases, Copenhagen University Hospital, Rigshospitalet, Blegdamsvej 9, DK2100 Copenhagen Ø, Denmark. E-mail: omland@dadlnet.dk

Accepted 9 July 2014

Abstract

The DHCS is a cohort of all HIV-infected individuals seen in one of the eight Danish HIV centres after 31 December 1994. Here we update the 2009 cohort profile emphasizing the development of the cohort. Every 12-24 months, DHCS is linked with the Danish Civil Registration System (CRS) in order to extract an age- and sex-matched comparison cohort from the general population, as well as cohorts of family members of the HIV-infected patients and of the comparison cohort. The combined cohort is linked with CRS, the Danish Cancer Registry, the Danish National Hospital Registry, the Danish Registry of Causes of Death, the Danish National Prescription Registry, the Attainment Register and the Integrated Database for Labour Market Research to get information on vital status, migration, cancer, hospital contacts, causes of death, dispensed prescriptions, education and employment. Using this design, rates of a range of outcomes have been compared between HIV-infected patients and the comparison cohort, as well as between families of these two cohorts in order to disaggregate the effects of HIV infection and familial/environmental factors. Data can be shared with foreign institutions following approval from the Danish Data Protection Agency. Potential collaborators can contact the study director, Niels Obel (e-mail: niels.obel@regionh.dk).

What is the rationale for the new focus and new data collection?

DHCS was initially established to monitor the HIV epidemic and effects of highly active antiretroviral therapy (HAART) in Denmark. The cohort comprised exclusively data on HIV-infected patients. This design allowed the authors to study the spread of HIV in Denmark and the impact on mortality and morbidity of characteristics such as demographics, HIV-related factors, initiation of HAART, and co-infections with hepatitis B and C. This design, however, did not allow for a comparison of rates of specific outcomes between HIV-infected patients and the background population. Therefore, a design including comparison cohorts from the general population was introduced in which up to 19 individuals with same sex and date of birth as the HIV-infected patients were extracted at random from national registries of the general population. Using this design, outcomes such as mortality,
neurocognitive disorders,\textsuperscript{3} cataract surgery,\textsuperscript{4} diabetes mellitus\textsuperscript{5} and cerebrovascular events\textsuperscript{6} were compared between HIV-infected patients and the comparison cohort. Due to the observational nature of these studies, the authors were not able to establish whether the associations demonstrated were causal or not. Therefore another design was applied, in which we also captured family members (parents and siblings) of the HIV-infected patients and of members of the comparison cohort (Figure 1).\textsuperscript{7–10} The basic idea of this design was to challenge the findings (associations) from comparing outcomes between HIV-infected patients and the general population. According to this design, two relative risks (RRs) were compared: RR\textsubscript{HIV vs control}: the RR found when comparing HIV-infected patients and the comparison cohort and RR\textsubscript{family}: the association found when comparing family members of HIV-infected patients with family members of the comparison cohort. An RR\textsubscript{HIV vs control} > 1 indicates an increased risk of the particular outcome in HIV-infected patients compared with the general population. If RR\textsubscript{HIV vs control} > 1, RR\textsubscript{family} > 1 and RR\textsubscript{HIV vs control} \approx RR\textsubscript{family} this indicates that the increased risk of the particular outcome observed in the HIV-infected patients might be due to familial or environmental factors, and not to HIV per se. If, however, RR\textsubscript{HIV vs control} > 1, and RR\textsubscript{family} < RR\textsubscript{HIV vs control}, HIV might explain some, if not all of the increased risk of the particular outcome. For instance, two studies from DHCS demonstrated an increased mortality among HIV-infected patients compared with the background population, especially in the pre-HAART era (RR\textsubscript{HIV vs control}).\textsuperscript{2,11} In a study of mortality in siblings of HIV-infected patients compared with siblings of the comparison cohort (RR\textsubscript{family}), the same excess in mortality was not observed, indicating that the contribution of familial or environmental factors to mortality in HIV-infected patients was of little importance.\textsuperscript{7} In two other studies, it was demonstrated that the incidence of lung, head and neck cancer was increased in HIV-infected patients compared with the general population.\textsuperscript{9,10} The risk of these cancers were also increased among parents of HIV-infected patients, indicating that HIV-infection might be a marker of behavioural or family-related risk factors for lung, head and neck cancer.\textsuperscript{9,10} Outcomes such as myocardial infarction and lung cancer have also been studied using this design.\textsuperscript{8,10}

Figure 1. Illustration of the principal matching and data extraction procedure, which takes place every or every second year on DHCS data. Initially, a matching procedure is performed using the Danish Civil Registration System. In this way, for each HIV-infected patient in the DHCS, 19 individuals from the general population without a HIV diagnosis are matched on age and sex. Then, for both HIV-infected patients and the control subjects, parents and siblings are identified. Data are then extracted from the national registries for all cohorts. Using these cohorts, two comparisons can be made; RR\textsubscript{HIV vs control} is the incidence of an outcome among HIV-infected patients compared with that of the comparison cohorts, whereas RR\textsubscript{family} is a comparison between families of HIV-infected patients and families of the comparison cohort. The idea behind this design is that if RR\textsubscript{HIV vs control} \geq RR\textsubscript{family}, then familial or environmental factors possibly explain a substantial part of the presumed effect of HIV, whereas if RR\textsubscript{HIV vs control} < RR\textsubscript{family}, HIV infection is more likely to be the cause of RR\textsubscript{HIV vs control}.
What will be the new areas of research?

The DHCS will continue to examine factors associated with prognosis among HIV-infected patients. However, according to the design discussed in the previous section, we will continue to compare the rate of different outcomes between HIV-infected patients and the comparison cohort (RRHIV vs control in Figure 1) and to examine whether familial or environmental factors (RRfamily in Figure 1) might explain the presumed effect of HIV-infection. Using this design, we have already studied mortality, cancer incidence and incidence of myocardial infarction, and future areas to examine are incidences of kidney disease, diabetes and change in socioeconomic status. However, in principle any outcome registered in a Danish nationwide registry can be examined using this design.

Who is in the cohort?

DHCS is an open cohort with continuous enrolment including all HIV-infected individuals aged 16 years or older at time of HIV diagnosis, seen in one of the eight Danish HIV centres after 31 December 1994. HIV-infected patients are identified by clinical staff at the eight participating centres and continuously enrolled in DHCS. Once a year, electronic laboratory reports of CD4 cell counts and HIV-RNA viral loads are merged with DHCS. During this procedure we also check for cases of missed HIV diagnoses; the risk that a patient is missed in DHCS is therefore negligible, once he or she has been seen in one of these eight centres.1 Compared with the 2009 Cohort Profile, baseline characteristics are essentially the same but more patients have entered the cohort, which in 2013 included 6434 HIV-infected patients, of whom 1369 had died during follow-up. Every 12-24 months the following data extraction procedure is executed by Research Services, Statens Serum Institut (Figure 1).12 First, for each DHCS member, all individuals from the Danish Civil Registration System are identified who: match the HIV-infected individual on sex and date of birth; are alive and living in Denmark at the index date of that HIV-infected patient (index date is the latest of date of HIV diagnosis or 1 January 1995); and are not members of DHCS. From this population, 19 individuals are extracted at random to constitute the comparison cohort. These individuals are not allowed to contribute to the comparison cohort more than once.

Also from the Danish Civil Registration System, family members (parents and siblings) of HIV-infected patients and the comparison cohort are captured. Subsequently, the complete study cohort, now including HIV-infected patients, the comparison cohort and family members of these two groups, is linked with national registries using the unique civil registration number in order to obtain data from the Danish Cancer Registry, the Danish National Hospital Registry, the Danish Registry of Causes of Death, the Danish National Prescription Registry, the Attainment Register and the Integrated Database for Labour Market Research. For certain purposes, associations are studied in subgroups defined by patient characteristics, such as a particular place of birth. The high number of controls per HIV-infected patients (19) in the original dataset allows us to form smaller comparison cohorts defined by such characteristics, for instance a comparison cohort consisting of eight individuals born in Denmark per HIV-infected patient (born in Denmark).13

The data linkage is subject to approval by the Danish Data Protection Agency, but not to an ethical permission from an institutional review board or to approval by members of the cohort.

To study the impact of specific characteristics not registered in national registers (such as tobacco use), comparison cohorts have also been constructed from the Copenhagen General Population Study, which is a prospective study of a cohort of individuals randomly selected from Greater Copenhagen.14-16 The Copenhagen General Population Study has been approved by a Danish ethical committee and individuals in the cohort have provided written informed consent.

What has been measured?

Since the 2009 Cohort Profile,1 the main change in study design has been the linkage with additional national registries as well as the systematic use of family and comparison cohorts as described above. Access to data from the Danish Civil Registration System enables us to capture the comparison cohort as well as ascertain immigration, emigration and death.17 From the Danish National Hospital Registry we obtain data on all hospital admission and discharge diagnoses since 1977 and all hospital outpatient visits since 1995.18 The Danish Cancer Registry has recorded data on cancer diagnoses in Denmark since 1943.19 Notification to the Danish Cancer Registry is mandatory upon a diagnosis of cancer, and unreported cases are identified through computerized links to the death certificate file, the Danish National Patient Registry and the Danish Pathology Data Bank. Comprehensive assessment has found the Danish Cancer Registry to be 95–98% complete and valid.19 For causes of death, we used the Danish Registry of Causes of Death, which contains information from all Danish death certificates issued since 1943.20 Whenever a Danish resident dies, the attending physician must report the cause of death and the chain of events leading to death can be described by specifying up to four diagnoses.20 Finally, DHCS data have been linked with
national registries run by Statistics Denmark. Contrary to the above-mentioned registries, data from Statistics Denmark are anonymized and accessed through a remote connection to a server at Statistics Denmark.21 We used the Danish National Prescription Registry, which was established in 1994 by the Danish Medicines Agency at the National Board of Health and records individual-level data on all redeemed prescriptions dispensed at Danish community pharmacies, with complete data since 1 January 1995.22 The register includes variables related to the drug user, prescriber and pharmacy as well as date of dispensing, product code and name, Anatomical Therapeutic Chemical Classification (ATC) code, dosage form and strength. We also have gained access to the Attainment Register, which was established in 1981, covers the entire Danish population and includes data on successfully completed educational attainments.23 We accessed the Integrated Database for Labour Market Research, which was established in 1980 and contains data on occupational status, welfare services, social benefits and primary source of personal income on all individuals with a Danish citizenship. Data are collected from the Danish tax authorities, educational institutions and employment services, and are administered by Statistics Denmark.24 In the Copenhagen General Population Study, study participants are interviewed about lifestyle and health-related factors, including smoking.

What has it found? Key findings and publications

In a highly cited paper from DHCS from 2007, it was estimated that a 25-year-old HIV-infected patient can expect to survive until the age of 68 in the HAART era, which was far more than in the pre-HAART era, but less than among the general population.2 The study also demonstrated that survival in HIV-infected patients without comorbidity, alcohol and drug abuse or HIV risk factors (defined as a detectable viral load, a CD4 cell count below 200 cells/µl or an AIDS-defining event) was the same as in the general population.25 The substantially improved prognosis of HIV-infected patients has led to a change of focus from an exclusive examination of HIV symptoms and HIV treatment. Using the design described above, authors from DHCS have examined in a range of studies whether associations between HIV infection and the rate of different outcomes could be explained partly by familial or environmental factors.

Beside the studies of lung, head and neck cancer discussed above, DHCS has also demonstrated a nearly 3-fold increase in mortality in HIV/HCV-coinfected patients compared with HIV-infected patients not coinfected with HCV.26 Interestingly, nearly the same increase in mortality was observed when comparing mortality in siblings of HIV/HCV-coinfected patients with that of siblings of HIV-monoinfected patients.7 These results indicated that the contribution of familial or environmental factors to mortality in relation to HCV infection might be of great importance.

In a study of cerebrovascular events, it was demonstrated that HIV-infected patients carry an increased risk compared with the general population.6 The same association was not demonstrated among parents of HIV-infected patients except those of HIV-infected patients infected through injection drug use.6 Therefore, the increased risk of cerebrovascular events could not be explained by familial or environmental factors. The incidence of ischaemic heart disease is increased in HIV-infected patients, especially after commencing HAART.27 Among mothers of HIV-infected patients, an increased risk of myocardial infarction was observed, as well as among fathers whose offspring was infected with HIV through injection drug use.8 It was concluded that family-related factors might contribute to the risk of myocardial infarction in HIV-infected patients.

As described above, comparison cohorts have been extracted from the Copenhagen General Population Study for various purposes. In one of these studies, it was demonstrated that HIV-infected smokers lose more life-years to smoking than to HIV.28 The excess mortality of smokers is tripled in HIV-infected individuals and the population-attributable risk of death associated with smoking is doubled among HIV-infected patients in Denmark compared with the background population. In the same study design, it has been demonstrated that an increased frequency of smokers in the HIV population explains most of the excess non-AIDS-defining cancer risk seen in the Danish HIV population.29 Recently we have made use of a new data source, namely Statistics Denmark, which governs data on education, occupation and income. When comparing HIV- and non-HIV-infected populations, we have demonstrated that low educational attainment, defined as mandatory education of up to 9 years, is associated with substantially and equipotent increased mortality related to alcohol use and smoking. This indicates that the increased mortality among HIV-infected patients with low educational attainment might stem from factors not related to HIV infection.30 In another study on employment, we demonstrated that employment rates increased greatly among HIV-infected individuals from 1996 to 2011, but remained lower than in the background population.13 This study also indicated that loss of employment in HIV-infected patients might be permanent, highlighting the need to improve association of HIV-infected patients with the labour market in the initial
phase of the disease. So far, it has not been examined whether associations between education, occupation and income might be partly explained by familial factors, but these questions are among future projects.

**What are the main strengths and weaknesses?**

The principal strength of DHCS is the high quality of data registered in DHCS as well as in the national registers with which DHCS data have been merged.\(^\text{14,17–19,22}\) Strengths also include the matched cohort design discussed above and illustrated in **Figure 1**. Regarding comparison cohorts constructed from national registers, these are genuinely representative of the background population as the individuals are extracted at random from databases of the complete Danish population. The unique combination of matched cohort design with accompanying familial cohorts and high quality register data enables DHCS to address questions that could hardly be answered in other settings. Some characteristics such as smoking are not registered in national population-based registers, which is why other data sources such as the Copenhagen General Population Study have been used to construct comparison cohorts. As individuals in the Copenhagen General Population Study are only recruited from the capital of Denmark and their inclusion requires informed consent, this cohort may not be representative of the general Danish population. The inability to examine whether familial risk factors are caused by genetic or behavioural factors is a limitation of the study design.

**Can I get hold of the data? Where can I find out more?**

Data can be shared with foreign institutions provided approval is obtained from the Danish Data Protection Agency. Potential collaborators can contact the study director, Niels Obel (e-mail: niels.obel@regionh.dk).

**Funding**

The DHCS has received research funding from The Health Foundation, the AIDS Fund, the Augustinus Fund, the Novo Nordisk Foundation, Preben and Anna Simonsen Fund, Roche, Bristol-Myers Squibb, Merck Sharp & Dohme, GlaxoSmithKline, Abbott, Boehringer Ingelheim, Janssen-Cilag and Swedish Orphan.

**Conflict of interest:** None declared.

**References**