

## Kommentar

af Ruth Emerek og Lisbeth B. Knudsen

# Brug dog registrene - der er behov for sociologer, der kan og vil!

Da Grundforskningsfonden indkaldte forslag til „oprettelse af forskningscentre for demografi“ (Danmarks Grundforskningsfond 1996), var det ud fra en erkendelse af, at befolkningsudviklingen „er af helt afgørende betydning for samfundsudviklingen på længere sigt“. Samtidig pointerede fonden, at „der i Danmark er et enestående og omfattende datamateriale, herunder både historisk materiale ... og nutidige databaser om for eksempel fertilitet, familiestrukturer, dødelighed, mobilitet, migration og sygdom“ (Grundforskningsfonden 1996:8).

Grundforskningsfonden havde allerede tidligere generelt forsøgt at fremme registerforskningen gennem etableringen af Center for Registerforskning i 1995<sup>1</sup>. Dette center så fonden selv som et led i, at „De enestående registre udnyttes ... til at skabe muligheder for, at Danmark på dette område kan placere sig blandt de toneangivende nationer i den internationale forskning“ (<http://www.dg.dk/aktiviteter.html>, februar 2002). Grundforskningsfondens ønske om at fremme forskningen i demografi-

ske problemstillinger ved hjælp af de unikke danske registre er således et eksempel på, at de efterhånden flere og mere omfattende registre bliver et stadig mere oplagt og påtrængende datagrundlag for forskerne. Det gælder også for de sociologiske forskere (Knudsen 1995).

Danske sociologer arbejder primært med tværsnitsanalyser, hvis de overhovedet arbejder kvantitativt. De har ikke qua deres uddannelse det rette metodiske værktøj til at anvende det unikke longitudinale potentiale i registrene - nemlig at man kan følge de enkelte individer over en årrække. Inden for andre fagområder har man imidlertid længe haft øjnene åbne for registrenes muligheder og baseret væsentlig forskning på registre, der dækker enten et stort udsnit af befolkningen eller hele befolkningen. Det gælder primært inden for medicin og økonomi.

De lægevidenskabelige forskere har med udgangspunkt i den registrering, der finder sted inden for sundhedssektoren, længe været aktive brugere af de offentlige registre. Emnerne har især væ-

ret behandlinger, sygdom, fødsler og dødsfald. Projekterne har ofte været tværfaglige og behandlet social- eller arbejdsmedicinske problemstillinger - områder, som en del sociologer også tidligt var involverede i (se blandt andet Dansk Selskab for Social Medicin (1977)). Da sundhedsvæsenets registre indeholder relativt begrænsede oplysninger om de enkelte personers sociale forhold, er de medicinske oplysninger i en del af disse analyser blevet suppleret med oplysninger om f.eks. erhvervsarbejde, uddannelsesniveau og familieforhold fra registre i Danmarks Statistik.

Økonomer har brugt registrene til at studere udviklingen i blandt andet arbejdsløshed, lønudvikling, iværksættelse og jobåbninger. Allerede først i 1980'erne blev der etableret en forløbsdatabase på 5 pct. af befolkningen ved Økonomisk Institut på Århus Universitet (Westergård-Nielsen 1984). De tidlige økonomiske analyser på disse og lignende registerdata blev hurtigt inddraget i – eller endog decideret lavet som led i – ministerielle udredninger og blev således en del af det politiske beslutningsgrundlag.

Nogle af de gennemførte registerbaserede undersøgelser har „smittet af“ på det generelle registerberedskab i Danmarks Statistik. F.eks. resulterede både den tidlige analyse af sammenhængen mellem erhverv og dødelighed (Lyng 1979) og den senere analyse af sygehusindlæggelser og sociale forhold (Steenen & Juel 1990) i dannelse af løbende opdaterede registre. Analyser af den erhvervs-specifikke dødelighed indgik senere i Middellevetidsudvalgets arbejde (Ingerslev, Madsen & Andersen 1994) og et opdateret registermateriale findes nu i Danmarks Statistik (Andersen, Laursen & Petersen 2001) ligesom der findes et årligt register over befolkningens sygehusforbrug, som indeholder oplysninger om indlæggelser fra Sundhedsstyrelsens

Landspatientregister og fra registre om uddannelse, erhverv, indkomst og familieforhold for alle de indlagte personer (Danmarks Statistik 2001). Derimod var det Danmarks halten-bagefter med demografiske analyser, bl.a. i forhold til de øvrige nordiske lande, der lå bag dannelsen af Fertilitetsdatabasen (Knudsen 1993).

De arbejdsøkonomiske studier inspirerede til dannelse af det første omfattende register for arbejdsmarkedsforskningen, IDA, der kobler individoplysninger om hele befolkningen fra blandt andet Befolkningsstatistikregistret, den Registerbaserede ArbejdsstyrkeStatistik (RAS) og Indkomststatistikregistret. Etableringen blev finansieret af det Samfundsvidenskabelige Forskningsråd og Danmarks Statistik i fællesskab og IDA blev placeret på Danmarks Statistik (Danmarks Statistik 1990).

Siden sidst i 1980'erne er det gået stærkt med etablering af nye databaser, og Danmarks Statistiks oversigt over forskningsdatabaser omfatter ikke færre end 23 databaser, der er mere eller mindre forskellige – men alle temmelig omfattende – og beregnet til forskellige formål<sup>2</sup> (Danmarks Statistik 2002). Nogle er finansieret af Danmarks Statistik – andre af forskningscentre eller forskningsinstitutioner. Hovedparten af disse databaser indeholder koblinger af data, som er yderst relevante i en sociologisk sammenhæng fx registret *Erhverv og cancer*, der kobler oplysninger fra *Cancerregistret* med oplysninger fra Danmarks Statistik og registret *Erhverv og dødelighed*, der igen kobler data fra blandt andet Befolkningsstatistikregistret, Arbejdsklassifikationsmodulet (AKM) og Uddannelsesstatistikken til oplysningerne om dødsfald. Fertilitetsdatabasen, hvis etablering blev finansieret af Danmarks Statistik og det Samfundsvidenskabelige Forskningsråd, er især rettet mod demografisk forsk-

ning af fertilitetsudviklingen og henvender sig derfor klart til sociologer med dette forskningsfelt (Knudsen 1993).

Det seneste initiativ i retning af øgning af registerforskningen kom med IT- og Forskningsministeriets rapport, „Registerforskning – Enestående danske muligheder“ (IT- og Forskningsministeriet, 2001). Denne rapport fremlagde en række konkrete forslag til, hvordan adgangen til registre kunne gores både bedre og billigere, således „at registerforskningen kan blive et flagskib indenfor dansk forskning“, som det smukt hedder i forordet til rapporten. Udvalget bag rapporten formulerede således, at Danmark internationalt set er „et foregangsland hvad angår mulighederne for at gennemføre registerforskning“ (ibid., p.5). Det kan vi jo kun være enige i. Der er således kræfter i gang for at fremme registerforskning – også gennem en sikring af at væsentlige forudsætninger for en mere omfattende brug af registre opfyldes, nemlig at de strukturelle begrænsninger for adgangen til registerdata fjernes, og at prisen for anvendelsen sænkes<sup>3</sup>.

Der er derfor mange argumenter for, at også sociologer skal tage fat på registre. De edb-tekniske muligheder for at håndtere de store datamængder i registre er stærkt forbedret både med hensyn til brugervenlighed og hurtighed og i forhold til, hvor komplekse analyser, der kan gennemføres. Longitudinelle studier på registre bør derfor være en naturlig og realistisk del af de metoder, sociologen kan vælge at anvende i sine analyser. Det kræver ganske vist, at de sociologiske miljøer i højere grad erkender og accepterer nødvendigheden af en flerhed af metoder – en flerhed, der også omfatter de kvantitative metoder, der bygger på en noget anderledes matematisk og statistisk tilgang, end den man som sociolog hidtil har tilegnet sig. Det er selvfølgelig ikke realistisk at opkvalificere

alle allerede færdiguddannede sociologer til habile „registerforskere“, men det er netop derfor bydende nødvendigt at sikre, at fremtidens sociologer *alle* er godt rustede til at vurdere, hvornår registerdata evt. kan være en væsentlig basis for en undersøgelse, og at *flere* reelt bliver i stand til at udføre sådanne analyser.

De kommende sociologer må derfor som minimum forsynes med metodiske redskaber, der sætter dem i stand til at bruge registre i deres analyser – eller til evt. at give registre et indforstået, kvalificeret modspil. Sociologerne bør ikke fortsat overlade det datamæssige forskningspotentiale, der ligger i registre, til andre faggrupper. Sociologi bør tværtimod i fremtiden tælles med blandt de fagområder, der har glæden af de ret så enestående danske registre.

## Hvad gør vi i undervisningen?

Sociologistudiet på Aalborg Universitet har valgt at klæde de studerende på både metode- og videnskabsmæssigt med hensyn til at anvende registre. Et obligatorisk kursus i *Registeranalyse* (under titlen: *Registre og forløbsanalyser*) er en del af den videregående metodeundervisning på 7. semester. Kurset bygger på et mindre, individbaseret, longitudinelt modeldatasæt, SOC-DATA, der er specielt fremstillet til undervisningsbrug på sociologiuddannelserne<sup>4</sup>. Dette datasæt består af et udtræk fra forskningsdatabasen IDA, og det indeholder oplysninger om et udtræk af personer, der følges hvert år gennem perioden 1980-1995, hvor de er i aldersgruppen 20-59 år. I alt indeholder modeldatasættet oplysninger om knap 3300 personer samt om deres eventuelle gifte/samboende partnere i perioden. Datasættet omfatter 23 sociologisk relevante variable, såsom uddannelse, erhvervs- og familieforhold, samt baggrundsvariable som køn og alder, og det giver mulighed for at gennemføre både tvær-

snits- og longitudinelle analyser på individniveau<sup>5</sup>. I løbet af kurset, som nu er blevet gennemført for de to første årgange sociologistuderende, skal de studerende ud fra en selvvalgt problemstilling gennemføre en longitudinal analyse<sup>6</sup>.

Det har været interessant at se, hvordan de studerende har brugt deres sociologiske fantasi til at undersøge mange forskellige problemstillinger på det trods alt begrænsede materiale. De har f.eks. analyseret hvad, der har betydning for, om man bliver arbejdsløs, får deltidsarbejde, gør lønmæssig karriere eller opnår forfremmelse på anden vis. Mange studerende har interesseret sig for overvejende demografiske problemstillinger, som f.eks. om – og hvornår – man bliver gift, med hvem, om man får børn, hvem der bliver skilt og hvad børnene betyder i denne sammenhæng. Andre studerende har undersøgt, hvem der flytter fra hovedstaden, og hvem der flytter til landkommunerne.

Nogle har arbejdet eksplorativt med data og har på eksemplarisk vis gradvis udbygget deres analyse ved at inddrage deres sociologiske viden om forventede sammenhænge. Men hovedparten af de studerende har forsøgt at belyse en eller flere sociologiske teorier (primært af Giddens, Sennett, Bordieu og Simmel) med deres empiriske analyse, samtidig med, at de ofte har kombineret dette med en demografisk tilgang og blandt andet anvendt kohorteanalyser. Generelt har det været indtrykket, at de studerende på dette relativt korte kursus har fået indsigt i registerforskningens muligheder og begrænsninger, og det er glædeligt at se, at registrene også overvejes som en del af datagrundlaget for de første specialer på sociologistudiet i Aalborg.

Det har også været dejligt at se sociologistuderende begynde at tænke i demografiske baner igen. Det inspirerer modeldatasættet selvfølgelig til alene

qua sit indhold, men en væsentlig årsag er uden tvivl, at sociologistudiet i Aalborg fra begyndelsen har omfattet et tilbud om Demografi som valgfag på bacheloruddannelsen. Kurset følges hvert år af ca. en tredjedel af årgangen. I undervisningen får de studerende viden om den demografiske udvikling og stifter bekendtskab med et redskab til at beskrive og analysere udviklingen. Hovedvægten er lagt på danske forhold, men med et internationalt perspektiv, bl.a. på vandringer, fertilitet, familiedannelse, dødelighed og befolkningsudvikling<sup>7</sup>.

## Perspektiver

Med de eksisterende registre, som løbende opdateres, kan man på de fleste områder belyse de sidste tyve års udvikling. Ved at bruge registrene slipper vi for recall-bias, vi slipper for de bortfald, der ikke skyldes udvandring eller død, og vi slipper for den manglende repræsentativitet, som efterhånden kan præge flere og flere spørgeskemaundersøgelser på grund af befolkningens stigende modvilje mod at svare på de stadig flere spørgeskemaer. Men vi slipper selvfølgelig ikke for væsentlige diskussioner om definitioner og operationaliseringer, og problemer med at tilskære data (der er indsamlet til nogle formål) til anvendelse til andre, nye formål, og vi kan ikke altid være sikre på, at de tilgængelige data akkurat passer til vores problemstillinger.

Behovet for både pålidelig og gyldig viden bliver ikke mindre, blot fordi befolkningen tilsyneladende bliver mere og mere træt af at deltage i interviews og besvare spørgeskemaer – tværtimod. Det er derfor nødvendigt at erstatte noget af den viden, som man i øjeblikket indhenter fra befolkningen gennem spørgeskemaer, med den viden som faktisk allerede eksisterer – måske i en anden udformning – i registrene. Dette er endnu et

argument for, at sociologerne skal søge andre og nye veje med hensyn til valg af deres empiriske forankring. Endelig har de sidste 20-30 års udvikling tydeliggjort nødvendigheden af, at sociologerne flytter fokus til dynamiske processer frem for statiske analyser: F.eks. har den befolkningsmæssige udvikling sat mere og mere fokus på nødvendigheden af at studere den demografiske udvikling ud fra en sociologisk vinkel. Det gælder bl.a. de dramatiske ændringer i fødselsrængeles størrelse, de tilbagevendende diskussioner om forsørgelsesgrundlaget for velfærdsstatens relativt mange gamle og den etnisk mindre homogene befolkning i Danmark.

Der er nok at tage fat på – og i Danmark har vi enestående muligheder for at gøre det.

## Noter

1. Samtidig blev der etableret en filial af Danmarks Statistik i Århus, en særlig main frame til forskningsprojekter i Danmarks Statistik (Forskningsmaskinen) og Enheden for Registrering og Arkivering af Sundhedsvidenskabelige data.
2. Hertil kommer det store antal registre, som hver dækker ét område og fra hvilke data kan udtrækkes til specifikke projekter.
3. Adgangen til registerdata har hidtil forudsat en forskerplacering på Danmarks Statistik i København eller (fra 1995) Danmarks Statistiks filial i Århus. Desuden har prisen for anvendelse af data for mange forskere været en hindring for at anvende registerdata. Anvendelsen har som oftest fordret en forskningsbevilling og en institutionel forankring. Regeringsskiftet i november 2001 satte dog en foreløbig stopper for de nævnte initiativer, der skulle lette anvendelsen, men i slutningen af marts 2002 er det blevet oplyst, at Danmarks Statistik har fået en bevilling, som vil resultere i prisnedsættelser for registerkørsler. Endnu i skrivende stund ved vi dog ikke, hvorledes dette udmøntes.
4. SOC-DATA modeldatasættet blev etableret i 2000 i et samarbejde mellem Gerda Engholm (Center for Registerforskning),

Søren Leth-Sørensen (Danmarks Statistik), Anders Holm (Sociologistudiet i København) og Ruth Emerek (Sociologistudiet i Aalborg).

5. Anonymiteten af data er sikret bl.a. gennem gruppering og scrambling af data. Derudover underskriver hver studerende en fortrolighedserklæring for at få adgang til datasættet.

6. En nærmere beskrivelse af kurset kan ses på Sociologistudiets hjemmeside, se fx <http://www.socsci.auc.dk/sociologi/Kursuskataloger/7-sem-e01/metode-registre-og-forl.htm>.

7. I forbindelse med dette kursus skal de studerende udarbejde en ugeopgave om et bundet emne – en nærmere beskrivelse af kurset kan ses på sociologistudiets hjemmeside, se <http://www.socsci.auc.dk/sociologi/Kursuskataloger/Valgfag-i-Demografi/demografi-f02.pdf>.

## Litteratur

- Andersen O., Laursen L., Petersen J.K. 2000: *Dødelighed og erhverv 1981-1995*. København: Danmarks Statistik.
- Danmarks Grundforskningsfond 1996: *Danmarks Grundforskningsfond indkalder forslag inden for Bioinformatik, Demografi, Geofag, Kemi, Matematik, Menneske-maskine grænsefladen, Plantebiologi*. København: Danmarks Grundforskningsfond.
- Danmarks Statistik 1990: *IDA – en integreret database for arbejdsmarkedsforskning, Hovedrapport* (ved Ruth Emerek, Per Vejrup Hansen og Søren Leth-Sørensen), København: Danmarks Statistik.
- Danmarks Statistik 1994: *Personstatistik i Danmark. Et registerbaseret statistiksystem*. København: Danmarks Statistik.
- Danmarks Statistik 2002: *Forskningsdatabaser – Databeredskab i Danmarks Statistik*. København: Danmarks Statistik.
- Danmarks Statistik 2001: „Sygehusbenyttelse 1999“ i *Statistiske Efterretninger*, serien Sociale forhold, sundhed og retsvæsen. 2001:5.(14. marts 2001).

- Dansk Selskab for Social Medicin 1977: *Registerbaseret social- og sygdomsstatistik: Arbejdsmiljø*. København: Dansk Selskab for Social Medicin.
- Ingerslev O., Madsen M., Andersen O. 1994: *Sociale forskelle i dødeligheden i Danmark*. 3. delrapport fra Sundhedsministeriets Middellevetidsudvalg.
- IT- og Forskningsministeriet 2001: *Registerforskning - Enestående danske muligheder*. København: IT- og Forskningsministeriet .
- Knudsen, L. B. 1993: *Fertility Trends in Denmark in the 1980s. A Register Based Socio-demographic Analysis of Fertility Trends*. København: Danmarks Statistik.
- Knudsen, L. B. 1995: „Kan registerforskning bruges i samfundsvidenskaberne?“ i *Dansk Sociologi* 1995; 6 (4): 96-99.
- Lyngge, E. 1979: *Dødelighed og erhverv 1970-1975*. København: Danmarks Statistik.
- Steensen, J.P. & Juel, K. 1990: *Sygehusindlæggelser og sociale forhold*. DSI-rapport 90.07. København: DSI/DIKE.
- Westergård-Nielsen, N.C. 1984: „Description of a Danish Longitudinal Data Base“, *Workingpaper 84-1*, Centre for Labour Economics, University of Aarhus and Aarhus School of Business.
- Ruth Emerek  
E-mail: emerek@i4.auc.dk
- Lisbeth B. Knudsen  
E-mail: lbk@demfo.sdu.dk