

Persondatabaser som kildemateriale for historiske undersøgelser

AF

NANNA FLOOR CLAUSEN & HANS CHR. JOHANSEN

I de senere år er der i en lang række lande blevet opbygget meget store databaser med personoplysninger, og der er på internettet givet adgang til at søge i baserne. Som oftest giver det kun mulighed for at finde frem til enkeltpersoner, men i mange tilfælde er det muligt fra værterne at få store datamængder leveret i en form, der tillader videre analyser baseret på massedata, enten i form af de samlede oplysninger eller som udtræk, der omfatter personkredse karakteriseret f.eks. ved et erhverv, en bestemt civilstand, tilhørende en bestemt aldersklasse eller afgrænset lokalt. Det åbner dels op for at kunne belyse helt nye historiske sammenhænge, dels for at få et bredere og dermed bedre dækkende grundlag for problemstillinger, som hidtil har været baseret på et langt mere spinkelt materiale. Men danske historikere synes indtil nu kun i meget begrænset omfang at have været opmærksom på disse nye muligheder.

Man kan undertiden møde den indvending, at da databasernes oplysninger ikke er den primære kilde, men kun en afskrift, bør man være varsom med at bruge dem. Det gælder dog i langt de fleste tilfælde, at der er gennemført en korrekturlæsning af materialet, og til de fleste formål vil en mindre fejlprocent kun have en meget begrænset indflydelse på et resultat, der er baseret på masseiagttagelser. Endvidere er de massekilder, det her drejer sig om, i sig selv også underlagt statistikens love, så behandlingen af »afskrifterne« er ikke væsensforskellig fra behandlingen af »originalerne«.

I det følgende vil der blive givet en oversigt over de større danske persondatabaser samt eksempler på udenlandske sidestykker, og i de efterfølgende afsnit vil der blive peget på de muligheder, der er for at udnytte materialet i historiske undersøgelser.

Eksisterende danske persondatabaser

De største danske datasamlinger af denne art er de offentlige registre, hvoraf det vigtigste er det centrale personregister (CPR). Da det indeholder en række personfølsomme data, er der ikke fri adgang til at benytte det, men Danmarks Statistik har oprettet en særlig forsknings-tilgang, hvor der er mulighed for at arbejde med dataene i anonymiseret form. Da det drejer sig om nutidige data, er det især inden for samfundsvidenskabelig og sundhedsvidenskabelig forskning, at denne mulighed er blevet udnyttet.

Den største danske persondatabase med historiske data er *Dansk Demografisk Database*, der er blevet til ved et samarbejde mellem repræsentanter for arkivvæsenet og amatørhistorikersammenslutninger. Der blev tidligt i projektføreløbet i 1990'erne udarbejdet et forlæg for indtastning af folketællingsoplysninger,¹ som derefter afleveres til Dansk Data Arkiv (DDA), der er vært for databasen og foretager den nødvendige bearbejdelse af dataene til den præsentation på internettet, der blev åbnet i 1996 (<http://ddd.dda.dk>). Til forskningsbrug udleveres folketællingerne på de vilkår, der gælder for øvrige udleveringer af datamaterialer fra Dansk Data Arkiv. Der er nu indtastet 14,5 mio. personoplysninger med komplet dækning af folketællingerne 1787, 1801, 1834, 1840, 1845, 1850 og 1885 (København). Desuden er folketællingen 1880 næsten færdig, mens dækning af de øvrige folketællinger frem til 1930 er af svingende omfang.

Samtidig med forlægget til folketællinger blev også publiceret et forlæg til kirkebøger. Dette er blevet væsentligt forbedret i 2007, og der kommer nu en stigende mængde af disse data til DDA, men den samlede dækning af denne kildes oplysninger er endnu relativt beskedent.

Et mere specielt sigte har den såkaldte *Udvandrerdatabase* (<http://www.emiarch.dk>), der som værter har Det danske Udvandrerarkiv og Aalborg Stadsarkiv. I den findes oplysninger om 394.000 personer, der i årene fra 1869 til 1908 udvandrede efter køb af en oversøisk billet gennem en dansk udvandringsagent. Oplysningerne omfatter ud over navnet på den emigrerede person tillige angivelse af, hvor i Danmark personen kom fra, hvor han eller hun skulle hen, samt alder og tidspunkt for udvandringen.

Flere andre persondatabaser er lokalt afgrænsede. Størst blandt dem er internetudgaven af *Politiets Registerblade* (<http://www.politietsregisterblade.dk>), der har Københavns Stadsarkiv som vært. Registerbladene, der blev indført i 1890, var en forgænger til de folkeregistre, der

¹ Denne vejledning betegnes SAKI-vejledningen og er publiceret som *DDA-Nyt*, nr. 65, 1993, pp. 7-137.

blev oprettet over hele landet i 1924. Bladene dækker for de mellem-liggende 34 år samtlige beboere over ti år i København, således at det enkelte blad er indrettet med de øvrige familiemedlemmer anført på husstandsoverhovedets kort. For hver person er opgivet navn, erhverv, fødested, fødselsdato og adresser i København i perioden. Det samlede antal personer i databasen er på 1,2 mio.

En lokal base med et bredere sigte er *Odensedatabasen* (<http://www.odensedatabasen.dk>), oprettet af Odense Stadsarkiv. Den har til formål at registrere byens befolkning i perioden 1740 til 1924. Den har et mere blandet kildemæssigt grundlag end de ovennævnte persondatabaser, idet den medtager folketællinger, kirkebøger, skattelister, borgerskaber, lavsoplysninger m.m. og gennem et personnummersystem kæder oplysningerne om den enkelte person sammen, så der opstår minibiografier af personens aktiviteter i den periode, han eller hun har opholdt sig i byen. Der er indtil nu indlæst ca. 300.000 personoplysninger i databasen, der fortsat er under udvikling.

Endelig bør nævnes, at The Church of Jesus Christ of Latter-day Saints (Mormonkirken) i Utah har oprettet en database kaldet *Family Search* (<https://familysearch.org/search/collection/igi>), der indeholder et meget stort antal personoplysninger fra hele verden, herunder et par millioner fødsler og vielser fra danske sogne. I den sammenhæng, der her behandles, er denne database dog mindre anvendelig. De medtagne oplysninger om den enkelte begivenhed er som oftest meget kortfattede. Ved fødsler er der i en betydelig del ikke anført fødselsdato, men kun dåbsdato, og ved vielser er der kun sjældent anført de viedes alder, selv om den normalt er angivet i kirkebogen efter 1814. Der er også kun opgivet navne, men ikke erhverv hos fædre eller de personer, der er blevet gift. Disse typer af oplysninger er særdeles vigtige ved de typer af undersøgelser, der er omtalt nedenfor. Family Search vil derfor især kunne bruges til at lokalisere enkeltpersoner, hvorefter man må gå til originalen for at supplere op med de manglende oplysninger.²

Eksempler på udenlandske persondatabaser (individ-/nominelle databaser)

I adskillige andre lande er der ligeledes blevet opbygget store persondatabaser og fortrinsvis med folketællinger som det centrale materiale,

² De originale folketællinger og kirkebøger er af Rigsarkivet affotograferet og lagt ud på internettet på adressen: www.arkivalieronline.dk. Dette og de øvrige netsteder, artiklen henviser til, er alle lokaliseret d. 19. august 2012.

men i flere lande, som f.eks. Holland, Sverige og Finland, blev de oprindelige skemaer kasseret, når de var blevet behandlet af de respektive statistiske institutioner, og man har derfor der måttet anvende andet materiale. I andre lande er de originale skemaer derimod blevet bevaret i det mindste fra nogle folketællinger. Det gælder f.eks. Storbritannien, Norge, Canada og USA. I det følgende gennemgås en række eksempler på persondatabaser, der er baseret på de originale folketællingsskemaer, og som primært dækker tællinger fra det 19. århundrede, samt nogle databaser, der anvender andet demografisk grundmateriale. Desuden gennemgås eksempler på internationalt samarbejde baseret på de nationale persondatabaser.

Norge

Norge har på nuværende tidspunkt (2011) folketællingerne for 1801, 1865, 1900 og 1910 i komplette databaser, der er til rådighed for forskningsbrug. 1801-tællingen er foretaget sammen med den danske 1801-tælling og er derfor helt sammenlignelig med den danske del; den var den første, der blev komplet tilgængelig i databaseform. Tællingerne for 1865 og frem kan findes hos *Registreringscentral for historiske data* (RHD), der er en national institution ved Universitetet i Tromsø. RHD's hovedformål er at dataregistrere historiske kilder og gøre disse tilgængelige for forskning. (<http://www.rhd.uit.no/index.html>). Disse data er blevet standardiserede, så de er klar til forskningsbrug og kan downloades for yderligere analyser. Der indtastes ved RHD desuden ejendomsregistre, kirkebøger og andre folketællinger end de nævnte.

Storbritannien

Storbritannien råder over den komplette folketælling for 1881, der omfatter England, Wales og Skotland og indeholder knap 30 mio. personer. Disse data er alle gennemgået og kodede til forskningsbrug. Det planlægges at få folketællingen for 1851 gjort komplet på samme vis. Der kan desuden søges efter enkeltpersoner i flere andre folketællinger. Disse data kan man få adgang til via NAPP-portalens: <http://www.nappdata.org/napp/>. Der er online søgeadgang til folketællinger for 1901 på <http://www.1901censusonline.com/main.asp?wci=welcome>.

Canada

Som Storbritannien råder Canada ligeledes over den komplette befolkning fra 1881 i databaseform: <http://www.prdh.umontreal.ca/census/en/main.aspx>. Der er 4,3 mio. personer i databasen. For folketællingen 1852 er de fleste skemaer for byerne kasseret efter den statistiske brug.

USA

For USA findes den samlede befolkning i 1880 i en database. USA er desuden langt fremme med at samle både yderligere komplette folketællinger og stikprøver fra folketællinger i en fælles database. Stikprøverne er fra de amerikanske folketællinger for 1850-2000. Der er til dette formål udarbejdet koder af alt indholdet for at standardisere det og gøre det muligt at analysere det. For at kunne anvende det udvælger man, hvilke oplysninger man vil analysere og for hvilke år, hvorefter de data, der opfylder kriterierne, kan downloades til analyse. De originale data for 1850-1950 er tilgængelige på mikrofilm på USA's National Archive, og det er desuden muligt at søge efter enkeltpersoner i USA i folketællingerne fra 1890 frem til 1930. Data er tilgængelige via Minnesota Population Center, <http://www.pop.umn.edu/>, der er vært for de amerikanske demografiske databaser.

Island

Island har den vistnok ældste statistiske folketælling fra 1703. Den kan findes i en database sammen med folketællingerne fra 1801, 1880 og 1901, indeholdende i alt ca. 140.000 personer. Derudover har Island en database, hvor alle islændinge fra en meget lang årrække er med, men den er ikke beregnet til historisk statistisk forskning. Der er adgang til de islandske data fra IPUMs (se afsnittet nedenfor om internationale samarbejdsprojekter) og NAPP (se afsnittet ovenfor om Storbritannien).

Sverige

Som nævnt ovenfor råder Sverige ikke direkte over bevarede folketællinger og afholdt i det hele taget ikke folketællinger som i Danmark før fra 1940. Sverige førte i stedet de såkaldte 'husförhörslängder', hvor

alle borgeres hændelser blev registreret, så längderne (*längd*: liste, fortegnelse) fik karakter af at være en slags folkeregister. Indtil 1855 blev befolkningstallet opgjort på baggrund af rapporter fra kirkebøgerne, hvorefter de blev erstattet af opgørelser fra 'husförhörslängderne' hvert 10. år. Disse opgørelser indeholder delvis de samme informationer, som vil kunne forventes i en folketælling. Ud fra dette materiale er folketællingerne fra 1880, 1890 og 1900 tilgængelige i en database. For de ældre perioder er husförhörslängderne noget ufuldstændigt bevaret. Der findes desuden i Sverige flere databaser, der ikke er landsomfattende og er baseret på forskellige typer af personoplysninger. De vigtigste er samlet på *Demografiska Databasen* i Umeå (<http://www.ddb.umu.se/databaser/>) med sammenkædede oplysninger fra kirkebøger og husförhörslängder i udvalgte områder, *Stockholms Stadsarkivs databaser* (<http://www.ssa.stockholm.se/Anvand-arkiven/Vara-arkiv/>), hvoraf den vigtigste indeholder oplysninger fra det såkaldte »Rotemansarkiv« om Stockholms borgere i perioden 1878-1926 (<http://www.ssa.stockholm.se/Anvand-arkiven/Folkbokforing-och-mantalskrivning/Rotemansarkivet/>), samt *Demografisk Databas Södra Sverige* ([http://ddss.nu/\(S\(nsu3i55ufsrhr45gf22ke55\)\)/swedish/default.aspx](http://ddss.nu/(S(nsu3i55ufsrhr45gf22ke55))/swedish/default.aspx)), med oplysninger fra kirkebøger i Skåne, Halland og Blekinge.

Internationale samarbejdsprojekter

De ovenfor nævnte persondatabaser er alle meget værdifulde, men deres værdi øges endnu mere, når de indgår i et samarbejde med andre. Fra 1990'erne og frem er der foregået et stadigt arbejde med dels at skabe de omtalte persondatabaser og derpå opbygge strukturer, der gør det muligt at udnytte det store indhold til at skabe ny forskning. I begyndelsen var samarbejdet fokuseret på at udpege fælles indhold og afklare, hvordan sprogforskelle kunne overvindes. På nuværende tidspunkt er samarbejdet kommet så langt, at forskere har direkte adgang til folketællingsdatabaser fra mange lande. De største internationale samarbejdsprojekter er:

NAPP: *North Atlantic Population Project*: www.nappdata.org/napp

IPUMS: *Integrated Public Use Microdata Series*: <http://www.ipums.org>

EHPS: *European Historical Population Samples Network*: <http://www.iisg.nl/hsn/news/ehps-net.php>

NAPP: Dette samarbejde har som forudsætning, at der er skabt databaser med historiske individdata. Projektet havde som udgangspunkt følgende medlemslande: USA, Canada, England, Norge og Island. Siden er Sverige kommet til, og i den kommende projektperiode vil der komme komplette folketællinger fra Danmark og Irland og stikprøver fra en lang række andre lande, bl.a Tyskland (Mecklenburg-Schwerin) og Egypten. De eksisterende data er alle blevet standardiserede efter en fælles udviklet struktur og meget nøje dokumenteret. Det gør det muligt at lave komparative demografiske og økonomiske undersøgelser over både tid og sted. Samtidig arbejdes der på at knytte personer sammen over tid for at kunne lave longitudinelle undersøgelser.

IPUMS: Med udgangspunkt i den ovenfor omtalte amerikanske digitalisering af folketællinger skal data fra NAPP ifølge projektbeskrivelsen for 2011-2016 inkluderes i IPUMS og standardiseres med dette projekt. Det vil skabe helt nye muligheder for at lave analyser af befolkninger over lang tid, byggende på de faktiske persondata og ikke på summerede, statistiske tabeller, der udelader detaljer.

EHPS: Dette projekt adskiller sig fra de to foregående ved dels at være et rent netværksprojekt og dels at bygge på persondata, der er kædet sammen på individniveau eller familieniveau, dvs. eksisterende longitudinelle databaser over individers livsforløb. Samarbejdet består i at forsøge at udvikle et fælles format, der vil give adgang til de databaser, der indeholder disse data, et såkaldt Intermediate Data Structure (IDS)-format. Når de eksisterende databaser har fået konverteret deres data til det fælles portalformat, skal der for at kunne hente data udvikles dataudtrækningsprogrammer, der er målrettet den analyse, man vil foretage. Der vil skulle anvendes forskellige data for eksempelvis at analysere enten fertilitet eller migration. Disse programmer vil blive udviklet inden for projektet, men samtidig kan alle selv bidrage med programmer, som man kan publicere på samme vis som fagfællebedømte artikler. I projektet deltager 14 europæiske lande med Holland (International Institute of Social History) som projektleder.

Klargøring af materialer gennem standardisering og kodning

En forudsætning for at kunne udnytte persondatabaserne til historiske forskningsformål er i de fleste tilfælde, at dataene gøres standardiserede. Det betyder ikke, at data ændres, men at de forskellige stavemåder

og udtryk for samme forhold ikke kommer til at stå i vejen for analyser.³ Normalt kan det gøres ved, at de oprindelige oplysninger bevares, mens resultatet af standardiseringen tilføjes i nye felter. På den måde kan det indtastede sammenholdes med det standardiserede, og begge dele kan sammenholdes med den oprindelige kilde.

Som eksempler på oplysninger, der kan være brug for at standardisere, før et materiale er klar til analyse, gennemgås i det følgende behov af denne type ved undersøgelser baseret på dansk folketællingsmateriale. Det sker ud fra erfaringer indhøstet på Dansk Data Arkiv.

Kønnet har og især havde uendelig stor betydning for et livsforløb, og derfor er det vigtigt, at alle personer har tilknyttet denne oplysning. I de skemaer, der blev uddelt i samtiden til udfyldelse ved folketællingerne, optræder der først fra folketællingen 1870 en særlig rubrik for køn, men i DDA har man ved at se på navnefordeling, stilling i husstande eller erhverv ved langt de fleste personer kunnet tilføje denne oplysning. For 1801-tællingen drejer det sig kun om 101 poster ud af knap 1 mio. poster, hvor oplysningen ikke kan tilføjes, og for 1850 har alle fået tilføjet kønnet.

I *aldersfeltet* skulle man først fra 1901 anføre nøjagtig fødselsdag og -år. Frem til og med folketællingen i 1860 skulle angives en alder med det 'løbende år iberegnet'. Det vil sige, at en nyfødt skulle anføres med alderen 1 år (var i sit første leveår). I tællingerne fra 1870 til 1890 skulle anvendes den moderne måde at angive en alder med det fyldte år. Disse forskelle må man være klar over ved sammenligninger mellem aldersfordelingerne i folketællinger fra flere år. På trods af instrukserne kan der imidlertid stå andet i feltet, f.eks. '3 uger'. Alderen er i sådanne tilfælde af DDA omregnet til et regulært tal. For alder over 100 år er der kontrolleret særligt ved at se i kilden. Ved anvendelse af de ældre tællinger skal man også være opmærksom på, at der – navnlig for de ældre personer – er en tendens til at opgive runde aldre.

Civilstand er et felt, der normalt er hurtigt at standardisere, fordi der kun er få varianter på udfyldelsen af rubrikken. Der forekommer imidlertid i de oprindelige lister en del tilfælde, hvor rubrikken ikke er udfyldt. I mange tilfælde drejer det sig om børn og unge, så der uden videre kan indføres et »ugift«. I andre tilfælde kan civilstanden fremgå af stillingen i husstanden (husstandsoverhovedets hustru), men på grund af for få oplysninger er det ikke altid muligt at afgøre civilstanden, og det er da nødvendigt at indføre 'Uoplyst' som en kode.

3 I Metode & Data, 2006:92, <http://ddd.dda.dk/ddakatalog/mogd/md92.pdf>, giver Hans Jørgen Marker en detaljeret gennemgang af standardiseringen af folketællingen fra 1801.

Oplysninger om *stilling i husstanden* kan give to problemer. I folketællingerne fra 1834 til og med 1860 er denne oplysning opført sammen med personens erhverv. I 1850 hedder rubrikken således »Personernes titel, Embede, Forretning, Haandværk, Næringsvei, Stilling i Familien eller hvad de leve af«. Det betyder ofte, at hvis der er anført et erhverv, er der ingen stilling i husstanden, så man må søge at udlede denne af den samlede familiesammensætning. Det andet problem kan være, at fjernere familiemedlemmer end mand, kone og børn måske er anført som tjenestefolk, så man ikke umiddelbart kan se, om der er tale om en udvidet familie, dvs. en der tæller flere end kernefamiliens medlemmer.

Standardisering og kodning af *erhverv* er meget kompleks. Det er der mange grunde til. Den første er den ovennævnte sammenblanding i nogle tællinger med stilling i husstanden, men det gælder også oplysningen om 'titel', der oftest vil give information om en persons status (f.eks. justitsråd, men ikke dommer), mens 'hvad de leve af' betyder, at vi får kendskab til de personer, der er på aftægt, modtager understøttelse og lignende. Et andet problem er, at det ikke ud fra det oplyste kan ses, hvilket erhverv der bibringer den største indtægt, når der er anført flere erhverv. Der findes flere forskellige principper for kodning af dobbelterhverv. I Norge har Digitalarkivet kodet erhvervsoplysningerne under et, således at f.eks. 'husmand, fisker' ikke er det samme som 'fisker, husmand'. Ved den kodning af de danske folketællinger, som er foretaget i undersøgelser på DDA, er 'husmand, fisker' blevet opdelt i to erhverv, der ved analyser hver kan vægtes som 0,5 erhverv for den pågældende person. Hvis der var tre erhverv, så som 0,33 erhverv etc.

Det næste problem ved kodningen af erhverv er den store forskel i detaljeringsgraden. I nogle sogne er 'husmand' simpelthen angivet som 'husmand', hvor det i andre kan være 'husmand med jord', 'husmand med lidt jord' eller 'husmand uden jord'. Andre eksempler er 'hørspinder', 'uldspinder', 'spinder'. For analysen af erhverv vil det normalt af hensyn til overskueligheden være vigtigt at arbejde med færre kategorier, i dette tilfælde 'spinder', hvor det for andre analyser kan have interesse, hvilken type spinder det drejer sig om.

For at kunne kode erhverv er der i det internationale samarbejde blevet udarbejdet en tabel med standardiserede tekster for erhverv, der tillige giver mulighed for at referere til de enkelte personer. DDA's kodning af de danske folketællinger sker i nært samarbejde med det internationale projekt for kodning af historiske erhverv, HISCO – Web-Based Information System on the History of Work: <http://historyofwork.iisg.nl/index.php>. Dette system bygger på det arbejde, International Labour Office har udført med at udvikle en international

standard for klassificering af erhverv. De danske koder er mere detaljerede end HISCO-koderne, så afhængigt af forskningsprojekt kan man anvende den ene eller anden kode, idet de danske henviser til HISCO.

Opdelingen på husstande skal i princippet foretages samtidig med indtastningen af folketællingen, men kvaliteten af husstandsidentifikationen er stærkt varierende, og dette – sammenholdt med, at strukturen i folketællingerne ofte er uklar og vekslende – har gjort det til en længe-revarende proces, hvor man i DDA søger at standardisere de modtagne indtastninger. Husstandene er blevet opdelt i typer, f.eks. Familie, Institution eller Soldater. Der er i alt fem typer. Dernæst er inddelingen sket ved at se på stilling i husstanden, hvor en husstand hyppigt vil begynde med betegnelsen 'Husfader', der angiver husstandsoverhoved. Derudover er det nødvendigt at se på de øvrige personers stilling i husstanden for at undgå en fejl, hvis f.eks. 'hans kone' er anført som den første person i en husstand.

Forskningsmuligheder – stort potentiale for nye undersøgelser

De typer af oplysninger, der findes i databaserne, vil først og fremmest have interesse for historikere, der beskæftiger sig med emner inden for økonomisk og social historie samt med befolkningsudviklingen, men der er også oplysninger, der kan anvendes ved andre former for undersøgelser. Dels kan databaserne give et større datagrundlag, dels giver de muligheder for at foretage andre grupperinger af personerne end dem, der kan findes i den samtidige offentlige statistik. Dertil kommer, at det ofte vil være muligt at foretage sammenfletninger af forskellige kilder, så der ad den vej kan opnås ny indsigt i en række samfundsforhold.

Inden for økonomisk historie vil det ofte være erhvervsoplysningerne, der har den største interesse. Angivelserne i databasernes grundmateriale vil være mere specifikke end i den offentlige statistik, og inden for de enkelte erhverv vil det være muligt at foretage opdelinger efter f.eks. alder, køn, adresse, hvorefter man kan følge en række udviklingsforløb eller sammenligne forholdene forskellige steder på samme tid. I de sene folketællinger findes også arbejdsstedet opgivet, så man ad den vej kan finde frem til de beskæftigede på et lands større virksomheder.⁴

4 Sådanne undersøgelser har især været gennemført i USA, fordi flere af de amerikanske folketællinger også indeholder en række oplysninger om de optalte personers erhvervsforhold ud over blot at nævne erhvervet. I Danmark er det kun folketællingen i 1916, der har ekstra oplysninger, nemlig i form af opgivelse af indkomst.

Inden for socialhistorien er der et bredt spektrum af muligheder for anvendelse af databasernes oplysninger. Det er f.eks. i folketællingerne muligt at finde frem til alle, der modtog fattighjælp, og ofte kan man se noget om, hvordan hjælpen blev ydet – i form af omgangsforsørgelse, ved at fattige børn blev sat i pleje, eller at modtagerne var beboere af et fattighus eller en fattiggård. Også »ældreproblematikken« kan belyses. Hvor stor var »ældrebyrden«? Boede de gamle sammen med deres gifte børn eller i selvstændige husstande? Hvor mange ældre er opført som almissemodtagere? Den sociale lagdeling i et samfund kan i nogen grad ses ud fra erhvervsbetegnelserne – eller mere indirekte ud fra antallet af kvindelige tjenestefolk i husstanden, men ofte vil det være muligt at sammenflette folketællingsoplysninger i et samfund med skattelister eller bygningsregistreringer, så man ad den vej kan få yderligere sociale indikatorer, der f.eks. kan vise indkomstulighed og sociale kvarteropdelinger i et bysamfund. En del socialhistorikere har desuden været interesseret i at belyse husstandssammensætningen og forandringer i denne over tid. En anden mulighed for en social analyse er mulig, fordi det i den danske folketælling i 1921 skulle opgives, hvis man var arbejdsløs. Da tællingen skete under krisen efter den første verdenskrig, har man der en mulighed for på individbasis at finde frem til, hvem der blev ramt af krisen. Lignende oplysninger findes i flere udenlandske folketællinger.⁵

Befolkningshistorien – den historiske demografi – der ofte opfattes som en særdisciplin inden for socialhistorien, er det område, hvor de store persondatabaser først er blevet benyttet. Interessen har bl.a. været samlet om at belyse den demografiske transition – overgangen fra samfund med høje fødsels- og dødsrater til de langt lavere rater, vi kender i dag. Man har forsøgt at tidsfæste de forskellige trin i udviklingen – dødeligheden faldt først, mens det skete betydeligt senere for fødselshyppigheden – og at finde frem til, hvilke faktorer i samfundsudviklingen der kan have haft betydning for denne udvikling. Til belysning af disse forhold kan folketællingerne ikke stå alene. De centrale kilder er her de digitaliserede kirkebøgers oplysninger om fødsler og dødsfald, og også vielseslisterne er vigtige, nemlig med henblik på at undersøge, om de typiske vielsesaldrer for kvinder ændrede sig, så antallet af fødsler (i ægteskab) ad den vej blev påvirket. Et andet forhold, der har haft interesse

5 Et tidligt eksempel på en socialhistorisk anvendelse er Stephen Thernstrom: *Poverty and Progress*, New York 1964. Senere er en række af de socialhistoriske undersøgelser blevet en del af den retning, der er blevet kaldt *New Social History*, hvor kvantitative kilder har spillet en stor rolle. Se herom Wikipedias artikel »Social History«, der indeholder en fylldig litteraturliste: http://en.wikipedia.org/wiki/Social_history.

for mange historikere, har været vandringer, således indvandringen til byerne under industrialiseringen og ikke mindst den oversøiske udvandring mellem 1865 og 1914. Flere lande har databaser med oplysninger om de udvandrede, og i USA findes der dels databaser, der direkte har registreret indvandringen i New Yorks havn, dels kan man i de amerikanske folketællinger hente oplysninger om, hvor en person er født, og hvornår den pågældende er kommet til USA.

Litteraturen på dette område er meget omfangsrig og alene nogle eksempler fra den nordiske litteratur skal nævnes for at vise spændvidden i emner.

Norge var meget tidligt klar med komplette folketællinger, og der er derfor i Norge megen forskning baseret på disse folketællinger. Et tidligt eksempel var Sølvi Sogner: *Folkevekst og flytting*, Oslo 1979. Fra de senere år kan nævnes: Arne Solli: *Livsløb – familie – samfunn. Endring av familiestruktur i Norge på 1800 tallet*, Bergen 2003. Denne afhandling bygger på folketællingerne fra 1801, 1865 og 1900. Gunnar Thorvaldsen har udgivet flere artikler og bøger, der baseret på folketællingerne analyserer migration, f.eks.: *Migrasjon i Troms i annen halvdel av 1800-tallet. En kvantitativ analyse av folketellingene 1865, 1875 og 1900*, Tromsø 1995.

I Sverige har især data i *Demografiska databasen* i Umeå tidligt dannet baggrund for megen forskning. På databasens hjemmeside kan man se en lang liste over udgivet forskning, bl.a. en bibliografi på 68 sider: http://www.ddb.umu.se/digitalAssets/61/61241_bibliografi1970-2010.pdf.

Senere er der også fra Lunds gruppe for historisk demografisk forskning kommet en del befolkningshistorisk litteratur, se f.eks. Martin Dribe: *Leaving Home in a Peasant Society*, Lund 2000.⁶

I Danmark er mikromateriale anvendt i Hans Chr. Johansen: *Befolkningsudvikling og familiestruktur i det 18. århundrede*, Odense 1975, og i *Danish Population History 1600-1939*, Odense 2002. Folketællingsmateriale er anvendt af medarbejdere ved Dansk Data Arkiv, se publikationer af Hans Jørgen Marker og Nanna Floor Clausen på arkivets netside <http://ddd.dda.dk/forskning.htm>.

Andre typer af historiske undersøgelser, der med fordel har kunnet benytte sig af databasernes oplysninger, har set på udviklingen i brugen af personnavne over tid, helbredsforhold, som de er oplyst i folketællingernes særlige rubrik om handicap – eller som det hed i den danske 1860-folketælling: »Døvstumme, Døve, Blinde og Forstanden

6 Så tidligt som i 1970erne blev der desuden i Sverige gennemført et stort emigrationsprojekt, hvor en del af undersøgelserne var baseret på befolkningsmateriale på mikroniveau. Se om dette projekt: Harald Runblom and Hans Norman (eds.): *From Sweden to America*, Uppsala 1976.

berøvede« – mens der i andre tales om krøblinge.⁷ Sammenfletning med andre digitaliserede kilder kan give endnu flere muligheder. Før der blev indført hemmelig afstemning ved de danske rigsdagsvalg kunne der forlanges skriftlig afstemning, og den enkelte vælgers votum blev så indført i en protokol. Kombineres valgbøgerne med folketællingsoplysningerne om den enkelte vælger, kan det sige en del om partiernes vælgerkorps.

De her opregnede eksempler på brug af de store persondatabaser er dog kun nogle smagsprøver. Det er i høj grad op til den enkelte historikers fantasi at se de muligheder, der ligger i den lette adgang til de efterhånden mange data, der findes i let tilgængelig form.

⁷ Birgit Eggert og Peder Dam fra henholdsvis Nordisk Forskningsinstitut og SAXO-Instituttet, Københavns Universitet, har benyttet folketællingerne til at undersøge fornavne og deres geografiske udbredelse, f.eks. i »Folketællingerne fortæller«, der kan læses på http://ddd.dda.dk/FT_om_navne.pdf.