

Personstatistik
på
registergrundlag

DANMARKS STATISTIK

18 AUG. 1982
DANMARKS STATISTIK
BIBLIOTEKET

Pris: 26,00 kr. inkl. 22 % moms
Danmarks Statistiks trykkeri, København
Udkommet i maj 1982

FORORD

Den 18. april 1978 påbegyndtes i Danmarks Statistik et registerstatistisk seminar. Formålet var at skabe grundlag for, at erfaringerne med statistikproduktion baseret på administrative edb-registre kunne nyttiggøres, dels ved planlægningen af den fremtidige registerudvikling, dels ved planlægningen af nye registerstatistikprojekter.

Til seminaret indbød man dels personer med særlige erfaringer fra registerstatistiske opgaver, dels personer, der stod foran at skulle planlægge eller lede sådanne projekter.

Man søgte at gennemgå og samle de vigtigste generelle erfaringer på dette område. Problemerne i forbindelse med samkøring af data fra forskellige registersystemer, specielt inden for personstatistikområdet, fik naturligt en fremtrædende plads.

Vægten blev lagt på den metodemæssige side af sagen, mens de legale aspekter ikke var tiltænkt nogen central rolle. Forberedelsen af registerlovgivningen var på det tidspunkt i fuld gang, og snart efter vedtoges Lov om offentlige myndigheders registre (13) og Lov om private registre mv. (14). Det skulle vise sig, at dette fik stor betydning for seminarets forløb, idet seminar deltagerne og især arrangørerne måtte påtage sig nye opgaver i forbindelse med lovens indkøring og administration, hvilket gjorde det vanskeligt at afslutte seminaret; et andet forhold, der virkede i samme retning, var dette, at problemerne omkring opbygning af et samlet statistiksystem i 1978 forekom for uafklarede til, at rapportens sidste kapitler kunne skrives. Først i sommeren 1981 genoptoges arbejdet med at redigere og færdiggøre det omfattende dokumentationsmateriale, som var produceret under seminaret. Resultatet offentliggøres i denne rapport.

Som følge af den lange tid, der er forløbet siden seminarrets start, er en væsentlig del af de eksempler, der findes i rapporten, nogle år gamle. Selvom der siden er sket en betydelig metodemæssig udvikling indenfor registerstatistikken, har det ikke været overkommeligt at

forny dette materiale, men eksemplerne er stadig egnede til at belyse problemerne.

Rapporten er udarbejdet i samarbejde mellem seminarrets deltagere. Den afsluttende redigering er foretaget af konsulent Lars Thygesen.

Danmarks Statistik i februar 1982

N.V. Skak-Nielsen

/Poul Jensen

Indholdsfortegnelse

| | Side |
|---|------|
| Kap. 0. Indledning og sammenfatning | 8 |
| 0.1. Rapportens sigte | |
| 0.2. Rapportens opbygning | |
| Kap. 1. Om begrebet register | 10 |
| Kap. 2. Statistikregistre og administrative registre | 12 |
| Kap. 3. Etableringen af administrative registre | 13 |
| 3.1. Personregistrering | |
| 3.2. Ejendomsregistrering | |
| 3.3. Erhvervsregistrering | |
| Kap. 4. Registerbaseret statistikproduktion | 17 |
| 4.1. Udviklingen i hovedtræk | |
| 4.2. Registerstatistik eller traditionel dataindsamling? | |
| Kap. 5. Statistikkenes krav til registersystemerne | 21 |
| 5.1. Indholdsmæssige krav | |
| 5.2. Registret som statistisk model | |
| 5.3. Statistikproduktion og registerorganisation . | |
| 5.4. Forskellige typer af registerdata | |
| 5.4.1. Identifikationsdata | |
| 5.4.2. Tidsdata | |
| 5.4.3. Basisdata | |
| 5.4.4. Specialdata | |
| 5.4.5. Baggrundsdata | |
| 5.4.6. Notatdata | |
| 5.4.7. Statistikdata | |
| 5.5. Dokumentation af registerdata | |
| Kap. 6. Samkøring af registeroplysninger | 34 |
| 6.1. Objektgruppeafgrænsning | |
| 6.2. Variabelkombination | |
| 6.2.1. Supplering af datasæt med ny variabel fra et andet | |
| 6.2.2. Sammenligning af variabelværdier | |
| 6.2.3. Konstatering af ændring i variabelværdi | |
| 6.2.4. Udvidelse af værdisættet i en given variabel | |
| 6.2.5. Ajourføring | |
| 6.3. Objektsyntese | |

| | Side |
|--|------|
| Kap. 7. Et registerstatistisk system | 41 |
| 7.1. Danmarks Statistiks målsætning | |
| 7.2. Datagrundlaget | |
| 7.3. Systemets elementer | |
| 7.4. Klassifikationsmodulerne | |
| 7.4.1. Karakteristik af klassifikationsmoduler | |
| 7.4.2. Krav til klassifikationsmoduler | |
| 7.4.3. Fordele ved at anvende klassifikations- | |
| moduler | |
| 7.5. Oprettelse og ajourføring af statistikregi- | |
| stre | |
| 7.6. Strukturen i det samlede statistikregister- | |
| system | |
| 7.7. Statistisk anvendelse af eksterne personop- | |
| lysninger | |
| Kap. 8. Fremtidig metodeudvikling | 54 |
| 8.1. Helt eller delvis manglende oplysninger | |
| 8.2. Supplerende dataindsamling: | |
| Estimation og kvalitetskontrol | |
| 8.3. Forløbsanalyser | |
| 8.4. Dokumentation af statistikregistre | |
| Bilag 1: Nogle infologiske grundbegreber | 61 |
| Bilag 2: 13 eksempler på samkøring | 63 |
| 1. Samlet anvendelse af børnetilskudsregistrene | |
| 2. En analyse af bruttobevægelserne i indkomsten | |
| fra 1970 til 1971 betinget af den beskæfti- | |
| gelsesmæssige indplacering | |
| 3. Uddannelses- og indkomstoplysninger sammen- | |
| holdt på udsnitsbasis | |
| 4. FTL-pilot 75 (dannelse af arbejdsklassifika- | |
| tionsregister) | |
| 5. Undersøgelse af sikkerheden i beregnede brut- | |
| toerhvervsforskydninger | |
| 6. Samkøring af viede til vielser | |
| 7. Uddannelsesstatistikken af 1974 | |
| 8. Statistik over døde fra 1973 | |
| 9. Statistik over levendefødte fra 1973 | |
| 10. Oplysninger om døde 1970-72 | |
| 11. Arbejdsstyrketællingen 1975 | |
| 12. Analyse af dødeligheden efter erhverv | |

| | Side |
|---|------|
| 13. Dannelse af oplysning om udlejningsforhold i bygnings- og boligregistret | |
| Bilag 3: Statistikregistre på personstatistikområdet og deres kilder | 115 |
| Bilag 4: De enkelte statistikregistre på personstatistik- området | 117 |
| Bilag 5: CPR's anvendelse i befolkningsstatistikken | 149 |
| Bilag 6: Klassifikationsmoduler i Danmarks Statistik | 157 |
| Bilag 7: Arbejdsklassifikationsmodulet (AKM) | 159 |
| Bilag 8: Milepæle i registerstatistikken siden 1968 | 167 |
| Bilag 9: Det erhvervsstatistiske registersystem | 169 |
| Bilag 10: Litteraturhenvisninger | 179 |

Kapitel 0. Indledning og sammenfatning

0.1. Rapportens sigte

Gennem 1970'erne skete en dybtgående omlægning af den samfundsbeskrivende statistik i Danmark: Mere og mere af statistikken er blevet baseret på administrative registre.

Denne rapport søger at samle de vigtigste af de erfaringer, der er gjort mht. udnyttelse af administrative registre. Rapporten omhandler navnlig det personstatistiske område, hvor etableringen af det centrale personregister (CPR) i 1968 og udviklingen af en række andre personorienterede, administrative edb-systemer allerede fra omkring 1970 skabte forudsætningerne for den registerstatistiske udvikling. Dog behandles også i nogen udstrækning personstatistikens berøringsflader med henholdsvis erhvervsstatistikken, hvor de registerstatistiske muligheder fra 1975 er baseret på det centrale erhvervsregister, og med boligstatistikken, hvis grundlag nu er det bygnings- og boligregister (BBR), som blev oprettet i 1977.

Rapportens sigte er først og fremmest at tjene som en grundig og generel indføring i problemerne for alle, der interesserer sig for registerstatistikken: Brugere af statistikken, nye ansatte i Danmarks Statistik, studerende m.fl. indenfor samfundsvidenskaberne. Desuden er det tanken, at erfaringerne mere målrettet kan nyttiggøres i kommende statistikprojekter i Danmarks Statistik og ved tilrettelæggelsen af den overordnede politik for statistikken udvikling i de kommende år.

0.2. Rapportens opbygning

Hovedvægten i rapporten er lagt på de metodemæssige spørgsmål inden for registerstatistikken. Derimod berøres andre vigtige spørgsmål såsom registerlovgivningen og de hertil knyttede spørgsmål kun perifert.

For at beskrivelsen af registerforhold kan blive præcis og give mulighed for at uddrage generelle konklusioner, benytter rapporten et begrebsapparat og en terminologi, som er stringent og logisk, nemlig den infologiske¹ model. Dette er et begrebsapparat, som er udviklet specielt til beskrivelse og analyse af registre og deres forhold til virkeligheden.

Hovedtrækkene i modellen er gennemgået i bilag 1.

De indledende kapitler beskæftiger sig med definitionen af begrebet register og afgrænser to hovedtyper af registre: Statistikregistre

¹Se fx Bo Sundgren (17).

og administrative registre. Dernæst gives i kapitel 3 et kort historisk rids af udviklingen af de administrative registre om personer, ejendomme og erhvervsvirksomheder. Registerstatistikens udvikling i årene efter CPR's oprettelse omtales i kapitel 4.

Herefter følger rapportens substantielle kapitler, først kapitel 5, som beskæftiger sig med de krav, de administrative registre bør kunne indfri for at muliggøre en effektiv statistikproduktion. De vigtigste krav er, at den type objekter, der omhandles i registret, afgrænses af en entydig definition, at hvert objekt kan identificeres entydigt, og at gruppen af objekter i registret er afgrænset af en regel. Det diskuteres udfra forskellige eksempler, i hvilket omfang disse krav kan ventes opfyldt i praksis, og hvad konsekvenserne af manglende opfyldelse kan blive.

Der introduceres en klassifikation af registerdata efter deres funktion i det administrative system, og de forskellige datatypers rolle i og værdi for statistikken belyses.

Samkøring af registeroplysninger fra forskellige kilder er en forudsætning for, at statistikken i større omfang kan drage nytte af de administrative registre. Analysen af samkøringserfaringerne indtager derfor naturligt en fremtrædende plads i rapporten. I kapitel 6 gennemgås en række forskellige typer af samkøring, som har været anvendt i registerstatistikken, og man diskuterer de perspektiver og vanskeligheder, de forskellige typer samkøring frembyder. Denne gennemgang er nøje knyttet til rapportens bilag 2, som beskriver og analyserer 13 eksempler på samkøringsprojekter i Danmarks Statistik, udvalgt med henblik på at belyse forskellige aspekter af samkøringsmetoderne.

Forudsætningerne for at fremstille statistik på basis af administrative registre har været gode i Danmark sammenlignet med andre lande. Fra begyndelsen af 1970'erne har Danmarks Statistik i sin strategiske planlægning sigtet mod at udvikle et registerstatistisk system, og konturerne af dette toner nu frem. Herom handler kapitel 7. Der gøres rede for elementerne i systemet og sammenhængen mellem dem. Et udtryk for systemets rækkevidde er det, at fremtidige folke- og bolig-tællinger foretages udelukkende på grundlag af de løbende registreringer, altså uden at borgerne skal udfylde folketællingsskemaer.

Imidlertid står mange problemer uløste tilbage, og der er behov for et omfattende metodeudviklingsarbejde fremover. Nogle af de felter, hvor behovene i dag træder stærkt frem, er omtalt i det afsluttende kapitel.

Kapitel 1. Om begrebet register

Betegnelsen register anvendes i forskellige betydninger inden for forskellige fagområder. Således omfatter registerbegrebet i henhold til almindelig edb-terminologi enhver logisk afgrænset samling af data, herunder også datasæt der er beregnet til rent midlertidigt brug; på den anden side opererer registerlovgivningen (8) med et begreb, hvor det afgørende er, om oplysningerne kan henføres til bestemte personer.

I nærværende fremstilling anvendes fortrinvis et snævrere registerbegreb. Ved et register skal herefter forstås en systematisk fortegnelse over en given gruppe af objekter¹ (fx personer, firmaer), som kan afgrænses præcist af en regel (fx "bopæl i Danmark"); for at man skal kunne tale om et egentligt register må det desuden kræves, at fortegnelsen og de informationer om objekterne, den indeholder, løbende skal ajourføres under hensyntagen til de ændringer, objektgruppen undergår. De følgende afsnit omhandler dog også datasamlinger, som ikke ajourføres på samme måde.

Et egentligt register må altså omfatte et system til indrapportering af dataændringer. I den forbindelse kan det nævnes, at ajourføringen kan finde sted med større eller mindre forsinkelse i forhold til den hændelse, der forårsager ændringen, og at indrapporteringen i nogle tilfælde ikke kan ske i tilslutning til hændelsen, således at man i stedet må nøjes med at konstatere de korrekte data med visse tidsintervaller; det gælder således nogle af de data, der findes i bygnings- og boligregistret, BBR. Data, der slet ikke kan ajourføres systematisk, betragtes ikke som egentlige registerdata.

Den anførte definition af registre støtter sig udelukkende til indholdet af de datasamlinger, der er tale om. Registrernes oplysninger kan opbevares på forskellig måde, et register kan eventuelt opdeles i flere delregistre af praktiske årsager, og der kan anvendes manuelle og/eller maskinelle processer ved vedligeholdelsen, men disse forhold spiller ingen rolle for registerbegrebets indhold.

Det må dog fremhæves, at de mere omfattende administrative registre, som i de seneste år har udgjort basis for statistikproduktionen, alle drives med anvendelse af elektronisk databehandling, og de erfaringer, som skal fremdrages her, vedrører derfor næsten udelukkende edb-registre.

Af den største interesse for statistikken er registre, som drives med en mangesidet anvendelse som formål, såkaldte basisregistre. Såle-

¹For en fremstilling af terminologien: Se bilag 1.

des må folkeregistersystemet og CPR i udpræget grad betegnes som et basisregister, idet formålet med dette system er at levere almindelige personoplysninger til brug overalt i den offentlige administration.

Et andet basisregister er erhvervsregistret, som registrerer grundoplysninger om selvstændige erhvervsdrivende og arbejdsgivere til brug for den offentlige forvaltning og andre. Tilsvarende må bygnings- og boligregistret betragtes som et basisregister.

De fleste registre oprettes imidlertid til brug for én bestemt forvaltningsopgave, fx skatteligningen.

Kapitel 2. Statistikregistre og administrative registre

Registre kan oparbejdes med forskellige formål for øje. En almindelig sondring går mellem administrative registre og statistikregistre. Denne sondring har stor betydning for afgrænsningen af de problemer, som behandles i nærværende fremstilling.

Som statistikregistre betegnes registre, der oprettes og vedligeholdes med det formål at frembringe statistiske opgørelser, dvs. optællinger, som ikke tillader identifikation af det enkelte objekt. Det er altså ikke hensigten, at oplysninger i disse registre om et objekt skal benyttes ved afgørelser i sager, der vedrører dette objekt. På den anden side kan statistikken benyttes ved overvejelser om udformning af regler, som har betydning for den enkelte.

Modsat står de administrative registre, hvis data indsamles som led i den administrative proces med det formål at danne grundlag for individuelle afgørelser angående den enkelte borger, familie, virksomhed osv.

Et statistikregister kan oprettes på grundlag af en traditionel statistisk data-indsamling, fx en folketælling eller en survey-undersøgelse, eller statistikregistret kan være bygget på kopier af eller uddrag fra administrative registre, som er underkastet en særlig statistisk bearbejdning. Det er hovedsagelig problemerne i forbindelse med statistik, der tilvejebringes på sidstnævnte måde, som skal behandles i denne rapport.

Det bemærkes, at registre, der udelukkende anvendes i statistisk eller videnskabeligt øjemed, i henhold til Lov om offentlige registre er underkastet særlige regler mht. egen-access for de registrerede og for videregivelse af oplysninger. For samkøring af registre til statistiske formål gælder ligeledes særlige regler.

Kapitel 3. Etableringen af administrative registre

3.1. Personregistrering

Det er indenfor personregistreringsområdet, at en bevidst udvikling hen imod basisregistre først kom i gang her i landet.

I løbet af 1800-tallet indførte mange vestlige lande en løbende og systematisk registrering af borgernes bopæls- og civilstandsforhold i såkaldte personregistre. Sådanne registre blev ikke dengang oprettet i Danmark, men i begyndelsen af det 20. århundrede opstod der som følge af væksten i dets offentliges administrative arbejde behov for pålidelige oplysninger om borgernes opholdssted. Således viste det sig vanskeligt for kommunerne at inddrive alimentationsbidrag og udskrevne skatter, fordi skatteyderne ikke kunne findes. Man blev også opmærksom på, at personregistre kunne tjene andre formål. Således ville de være egnede ved politiets eftersøgning af personer, som grundlag for udskrivning af valglistes, som grundlag for befolkningsstatistik mv.

I 1924 vedtog Folketinget derfor "Lov om Folkeregistre", som bestemte, at hver af landets kommuner samme år skulle oprette et folkeregister, dvs. et kartotek indeholdende oplysning om alle personer, der havde bopæl i kommunen, uanset om de måtte være midlertidigt fraværende.

Folkeregistrene skulle indeholde identifikationsoplysninger som stilling og navn, fødselsdag og -sted, og herudover var de vigtigste oplysninger bopælsadresse, familieforhold og statsborgerforhold.

Kommunerne skulle holde kartotekerne løbende ajourførte, idet de skulle modtage oplysning om fødsler, dødsfald, vielser, skilsmisser mm. fra forskellige myndigheder, medens borgerne selv blev forpligtet til at indberette flytninger samt ind- og udvandring direkte til folkeregistret.

I den debat, der gik forud for folkeregisterlovens vedtagelse, indgik også forslag om et landsregister, som skulle være en sammensmeltning af de kommunale folkeregistre omfattende hele Danmarks bosciddende befolkning, men dette forslag nød ikke fremme¹.

Folkeregistrene blev i de følgende år gradvist udbygget som hjælpemiddel for såvel den kommunale som den øvrige offentlige administration. Der blev indført formularsæt til bl.a. skattemyndigheder og sygekasser, således at disse myndigheder kunne føre egne personregistre. Efterhånden anvendtes hulkortteknik ved løsning af periodevise masseopgaver, såsom skatteberegning og udskrivning af valglistes. Til løs-

¹En nærmere omtale af baggrunden for folkeregistrenes indførelse og overvejelserne desangående findes i (2).

ning af disse opgaver oprettedes kommunalt ejede, regionale hulkortcentraler.

I 1968 gennemførtes en betydningsfuld reform af folkeregistersystemet, som havde til formål at forenkle og effektivisere den samlede offentlige personregistrering ved anvendelse af edb-teknik. Ved siden af de kommunale folkeregistre, som fortsatte med at fungere, oprettedes det centrale personregister, CPR, et landsdækkende magnetbåndregister over den danske befolkning.

En væsentlig del af reformen var indførelsen af et fast identifikationsnummer for hver enkelte borger, personnummeret. Dette nummer betragtedes som en praktisk nødvendighed for driften af det centrale personregister¹. Nummeret skulle tillige indføres overalt i den offentlige administration og således afløse de talrige nummersystemer, som tidligere havde været anvendt af forvaltningsgrenene.

Oprettelsen af CPR skyldtes først og fremmest ønsket om at undgå dobbeltregistrering og det ekstra ressourceforbrug, som fulgte heraf. Medvirkende har også været udsigten til en skattereform med indførelsen af kildeskat, som vanskeligt kunne gennemføres uden et meget sikkert system til identifikation af landets borgere.

I dag benyttes personnummeret som identifikation i en lang række specielle registre med personoplysninger, således:

Skatteregistre
 Pensionsregistre
 Arbejdsløshedsregistre
 Sociale registre
 Hospitalsregistre
 Registre over uddannelsessøgende
 Registre over værnepligtige
 Erhvervsregistret

De almindelige personoplysninger, som indgår i disse registre, vedligeholdes i vid udstrækning gennem automatiske meddelelser fra CPR.

CPR er, som det fremgår af denne beskrivelse, et virkeligt basisregister på personområdet, og de offentlige personregistres fælles anvendelse af CPR-data betyder en høj grad af koordinering af registrene.

Med hensyn til andre objekttyper end personer er udviklingen ikke nået helt så vidt, jf. de følgende afsnit.

¹Jf. (3).

3.2. Ejendomsregistrering

Ejendomsregistreringen her i landet kan siges at være startet med oprettelsen af matriklen i 1662, men dette register oprettedes udelukkende med sigte på jordbeskatningen. Oprindeligt indeholdt matriklen ikke opmåling af de registrerede arealer og deres grænser, men disse målinger blev snart foretaget og indført i registret. Dermed har matriklen fået en videre anvendelse, idet de registrerede grænser for jordejernes rådighed har betydning også i privatretlig sammenhæng. Senere har matriklen mistet sin helt direkte tilknytning til ejendomsbeskatningen, idet denne forvaltningsopgave siden begyndelsen af dette århundrede er blevet løst med udgangspunkt i særlige vurderingsregistre.

I 1972 oprettedes et landsdækkende edb-ejendomsregister til varetagelse af ejendomsbeskatning og ejendomsvurdering. Registret føres af kommunerne, og grundlaget for registreringen er oplysninger i matriklen og i tingbøgerne. Foruden matrikulære oplysninger mv. registreres forskellige data af betydning for vurderingen. I 1977 indsamledes i forbindelse med ejendomsvurderingen oplysninger til et nyt bygnings- og boligregister, BBR, som ligeledes føres kommunalt, og som rent teknisk udgør en udbygning af ejendomsregistret. Formålet med denne udbygning har været, at registerkomplekset skulle udgøre et egentligt basisregister for ejendomme, bygninger og boliger. De nye oplysninger omhandler bygningernes og boligernes alder, størrelse, installationer og anvendelse. Registret ajourholdes løbende med meddelelser om ændringer, nybygninger og nedrivninger, idet disse oplysninger indhentes i forbindelse med kommunernes byggesagsbehandling. Desuden foregår der en løbende kommunikation mellem BBR og CPR for at sikre, at adresserne i disse to basisregistre svarer til hinanden.

Oplysninger fra BBR indgår fra 1981 som basisoplysninger i grundlaget for de almindelige ejendomsvurderinger.

3.3. Erhvervsregistrering

Opbygningen i 1960'erne af administrative erhvervs- og virksomhedsregistre til brug for bl.a. ATP, toldvæsenet og skattevæsenet medførte indsamling af erhvervsdata, som også Danmarks Statistik kunne benytte til statistiske formål, herunder ajourføring af statistiske erhvervsregistre. De generelle erhvervstællinger, som Danmarks Statistik tidligere havde gennemført med tiårige intervaller, kunne på denne baggrund ophøre.

De administrative registres registrering af erhvervsvirksomhederne og arbejdsgiverne var imidlertid i den periode, registeropbygningen

fandt sted, kun i mindre omfang koordineret. Der var mange, men ofte små forskelle med hensyn til de forskellige registres enhedsdefinitioner, dataindhold mv. Endvidere anvendte hvert register sit specielle nummersystem til identifikation af enhederne.

Med henblik på en forenkling af det offentliges erhvervsregistrering gennemførtes i 1975 en lov om oprettelse af et centralt erhvervsregister.

Registret, der administreres af Danmarks Statistik, optager oplysninger om selvstændige erhvervsdrivende og arbejdsgivere, som i henhold til anden lov er berettiget eller forpligtet til registrering hos offentlige myndigheder eller institutioner. Registret optager endvidere oplysninger om filialer eller afdelinger under de oven for nævnte objekter.

Erhvervsregistrets oplysninger om firmaernes navn, adresse og indehavere registreres og ajourføres hovedsageligt ved direkte overførsel af oplysninger fra toldvæsenets og statsskattedirektoratets registre. Danmarks Statistik tilføjer oplysninger om brancher og særskilt beliggende afdelinger, filialer mv. (arbejdssteder) samt supplerende oplysninger om registerobjekterne.

I loven om erhvervsregistret forudsattes indført en entydig fælles nummerering af registerenhederne til brug i det offentliges erhvervsregistre.

Forslag til fælles registreringsregler og enhedsnumre er udarbejdet og i princippet tiltrådt af de myndigheder, der leverer oplysninger til erhvervsregistret. Fællesreglerne er imidlertid endnu ikke fuldt ud indført i erhvervsregistrets donorregistre. Erhvervsregistret har derfor endnu ikke kunnet etableres som en bestand af entydigt afgrænsede objekter fastholdt i et fælles identifikationsnummersystem. Dette har selvsagt vanskeliggjort anvendelsen af registret ved løsnin- gen af mange opgaver.

Kapitel 4. Registerbaseret statistikproduktion

4.1. Udviklingen i hovedtræk

Selv om folkeregistrene i 1924 blev oprettet med administrative opgaver som det helt klare hovedformål, var man fra begyndelsen klar over, at registrene tillige skulle anvendes til forbedring af befolkningsstatistikken. Folkeregistrene blev derfor pålagt at indsende flyttebeviser for samtlige flyttede personer hvert kvartal til Det statistiske Departement. Samtidig skulle sendes en særlig statistikindberetning, som udviste totaltal for antal personer i registret ved kvartalets begyndelse og slutning, hvortil kom en optælling af kvartalets befolkningsbevægelser af kategorierne: fødte, døde, indenlandske til- og fraflytninger samt vandringer til- og fra udlandet.

På grundlag af flyttebeviserne blev det muligt at fremstille en detaljeret statistik over flytningerne mellem kommuner og over ind- og udvandring. Statistikindberetningerne benyttedes til en summarisk opgørelse af "registerbefolkningen og dens bevægelser". Denne opgørelse måtte naturligvis være bundet af indberetningsformen og kunne fx ikke opdeles mht. alder og køn. Hovedparten af den detaljerede befolkningsstatistik hvilede fortsat på andet grundlag end folkeregistrene.

Oprettelsen af CPR i 1968 medførte en gennemgribende omlægning af statistikproduktionen. Det blev nu folkeregistrene, der via CPR kom til at danne basis for alle grene af befolkningsstatistikken. Nogle opgørelser henter fortsat oplysninger fra andre kilder, jf. eksempel 8 og 9 i bilag 2, men disse oplysninger opfattes som supplement til data fra CPR.

Indførelsen af den nye produktionsteknik gav i de første år anledning til betydelige vanskeligheder. Allerede i 1969 startede man forsøgsvis en befolkningsstatistik baseret på et CPR-materiale, som skulle beskrive befolkningen såvel som alle typer af befolkningsbevægelser. CPR-statistikken for 1969 og 1970 blev imidlertid særdeles mangelfuld, og det var først fra og med 1973, at det samlede statistiksystem kunne betragtes som oprettet.

Vanskelighederne skyldtes først og fremmest, at kommunikationen mellem statistikproducent og registermyndighed ikke var tilstrækkelig præcis. Som eksempel herpå skal anføres skilsmissestatistikken for 1970. Opgørelsen på grundlag af CPR gav som resultat, at skilsmisseantallet kun udgjorde en brøkdel af tallet for de foregående år, som var beregnet på grundlag af udskrifter af skilsmisseakterne. Årsagen viste sig at være, at man i ændringsmeddelelserne til Danmarks Statistik fra CPR ikke systematisk havde medtaget skilsmisserne, fordi af-

talerne på dette punkt ikke havde været udtømmende. Forud for den egentlige produktion af statistikken havde Danmarks Statistik gransket et udvalg af ændringsmeddelelser fra CPR, og da man i dette materiale havde fundet korrekte meddelelser om skilsmisser, troede man at grundlaget var i orden. Disse skilsmisser var imidlertid kun medtaget i materialet, fordi de fraskilte var flyttet, og den enkelte indberetning omfattede derfor flere (samtidige) hændelser.

Andre problemer skyldtes sammenblanding af forskellige datoangivelser for hændelser (flytninger, ind- og udvandringer).

Nogle af de data, som efter registerdokumentationen skulle indeholde statistisk operationel information, var i realiteten ikke udfyldt efter forskrifterne af alle kommuner. Af denne årsag var det i flere år ikke muligt at fremstille en fordeling af hele befolkningen på sogne.

Endelig skal omtales, at der i de årlige opgørelser af befolkningen og dens aldersstruktur forekom en fejl, som skyldtes en rent produktionsteknisk opdeling af registret ved ajourføring af CPR: Selve personregistret (CPR-masteren) blev ved enhver ajourføring kun opdateret med de ændringsmeddelelser, der var indkommet til den foregående ajourføring, medens de nye ændringer indgik i en "del-CPR-master", som alene omfattede de personer, der berørtes af ændringerne; denne del-CPR-master indgik først i selve CPR-masteren ved den næste ajourføring.

Den beskrevne produktionsmetode resulterede i, at det personregister, der blev dannet i en given ajourføring, fx uge 02, i virkeligheden kun var ført frem til uge 01. I Danmarks Statistik var man ikke bekendt hermed, og fejlen blev afdækket som følge af inkonsistens mellem forskellige dele af befolkningsstatistikken.

I de senere år har det befolkningsstatistiske system fungeret nogenlunde problemfrit. Samtidig har registerbaseret statistikproduktion fået voksende udbredelse. En oversigt over hovedpunkter i udviklingen siden 1968 ses i bilag 8.

Det er klart, at denne udvikling kun har været mulig som følge af den tidligere omtalte udbygning af det offentliges personregistrering og registrering af virksomheder, ejendomme mv.

Udvidelsen af registreringsaktiviteten er ofte sket udelukkende med administrative formål, men i andre tilfælde er hensynet til statistikproduktionen indgået med stor vægt i begrundelsen for oprettelse af nye registre eller udvidelse af bestående registres indhold. Det var således af stor betydning for vedtagelsen i 1976 af loven om det nye bygnings- og boligregister BBR, at registret kunne danne grundlag for en bolig- og beboerstatistik; i 1974 var det nemlig blevet beslut-

tet ikke at afholde den traditionelle folke- og bolig-tælling året efter under den forudsætning, at boligstatistik på lidt længere sigt kunne fremskaffes på registergrundlag. Ved opbygningen i 1960'erne af administrative erhvervs- og virksomhedsregistre indgik hensynet til anvendelse af oplysningerne til statistiske formål ligeledes i overvejelserne.

4.2. Registerstatistik eller traditionel dataindsamling?

Hvilken metode er optimal, når statistiske oplysninger skal tilvejebringes? Der er her tale om en kompleks problemstilling, og svaret afhænger naturligvis af de helt konkrete omstændigheder. Her skal blot fremføres nogle almindelige betragtninger.

En første forudsætning for en registerstatistisk udvikling er, at den er acceptabel for samfundet, dvs. at der ikke er opstillet love og regler, der hindrer en sådan udvikling. Denne forudsætning er og har altid været opfyldt her i landet.

Dernæst er det af betydning i hvilken udstrækning der rent faktisk findes etableret administrative registre med data af det ønskede indhold.

Hvis der ikke findes indholdsmæssigt dækkende registeroplysninger, bliver problemet, om statistikkens behov kan gøre en udbygning af registrene rimelig. Uden at komme nærmere ind på omkostningsspørgsmålet kan man sige, at nogle typer af statistik kun meget vanskeligt eller (i praksis) umuligt lader sig fremstille ved traditionelle metoder; der tænkes her på landsdækkende opgørelser, der angår den samlede befolkning, arbejdsstyrke, erhvervsaktivitet, boligbestand osv., og som skal kunne specificeres stærkt, fx i geografiske henseende. Specificationsgraden forudsætter en totaltælling eller i det mindste en meget omfattende stikprøveundersøgelse. Fx den lokale og regionale planlægning af investeringer og arealdispositioner vil oftest stille krav om statistik af denne art.

Når det drejer sig om statistik, der kan baseres på oplysninger om et repræsentativt udvalg af objekterne, således at indsamling og bearbejdning er overkommelig, vil valget ud fra økonomiske og kvalitative synspunkter derimod som oftest falde ud til fordel for de traditionelle statistiske metoder (postenquete eller interview), der bedre kan imødekomme mere specielle krav af indholdsmæssig karakter.

Valget mellem registerbaseret statistikproduktion og mere traditionelle metoder stiller sig anderledes, når det registermæssige grundlag i forvejen findes udviklet til administrative formål, idet udviklingsomkostninger til registersystemet ikke skal tages i betragtning

i denne situation. Under forudsætning af, at registersystemerne lever op til de indholdsmæssige krav, der beskrives i kapitel 5, vil det i almindelighed være fordelagtigt at basere statistikken på registrene. Der kan i mange tilfælde opnås en betydelig rationaliseringsgevinst, som samtidig gør det muligt at frembringe en rimeligt aktuel statistik.

Anvendelsen af registerbaseret produktionsteknik kan desuden tilføre statistikken en kvalitetsdimension, nemlig en indbyrdes konsistens mellem de forskellige statistikområder. Dette følger til en vis grad af, at registersystemerne indenfor fx personområdet i sig selv bliver bragt i harmoni gennem deres fælles anvendelse af basisregistret CPR. Herudover er det muligt i selve statistikproduktionen ved hjælp af sammenkoblingsteknik at overføre samme begreber fra en statistik til andre, jf. kapitel 6.

På den anden side må det fremhæves, at der naturligvis kan være kvalitative problemer med registerstatistikken. Det gælder med hensyn til statistikens relevans, kontinuitet og pålidelighed. Specielt kan der måske være en vis risiko for, at statistikken vil komme til at belyse virkeligheden alene fra de offentlige forvaltningsorganers synsvinkel. Det må være et krav, at sådanne problemer søges løst, hvilket ofte vil kræve betydelig indsats fra statistikerne.

Man kan næppe forestille sig, at alle kvalificerede statistikbehov skal kunne imødekommes gennem oplysninger, som findes registreret; fx vil statistik om holdninger i befolkningen ikke kunne fremskaffes ad denne vej. I mange tilfælde skal registerstatistikken derfor suppleres med traditionelle undersøgelser, som til gengæld kan baseres på repræsentative eller selektive udvalg. Også i sådanne undersøgelser kan det være hensigtsmæssigt at benytte registeroplysninger, nemlig som grundlag for udvælgelse af stikprøven. Det omtalte konsistenssynspunkt samt hensynet til de personer, som skal besvare evt. spørgeskemaer, tilsiger også i denne type af undersøgelser at overføre visse baggrundsoplysninger om de udvalgte objekter i stikprøven fra eksisterende registre. En særlig stilling indtager oplysninger, som kun kan konstateres i det administrative system, fx oplysninger om forvaltningens kontakt med klienterne. Ud fra et indsamlingsmæssigt synspunkt kunne disse oplysninger måske tænkes tilvejebragt uden brug af registre, men set fra et administrativt synspunkt vil det være fordelagtigt at indrette registersystemerne således, at de kan tilvejebringe de statistiske oplysninger som et bi-produkt. Dette synspunkt har været udslagsgivende ved tilrettelæggelsen af bibrandslovsstatistikken fra 1976 og arbejdsløshedsstatistikken fra 1979.

Kapitel 5. Statistikkens krav til registersystemerne

Dette kapitel omhandler de krav, som de grundlæggende registre bør kunne opfylde, hvis registerstatistikken skal have gode betingelser. Det er klart, at kravene ikke kan forventes opfyldt fuldt ud i praksis, og der foreligger en del erfaringer om konsekvenserne af sådanne mangler i registrene.

Indledningsvis kan statistikkens krav til registrene opdeles i indholdsmæssige og systemmæssige krav. Sidstnævnte omtales nærmere i afsnit 5.2-5.5, mens indholdskravene behandles i næste afsnit.

5.1. Indholdsmæssige krav

De indholdsmæssige krav er det vanskeligt at behandle generelt. Det drejer sig altså om at afgøre, hvilke oplysninger der skal findes registreret, for at man på grundlag af registrene kan få en "rimelig" dækning af statistiske behov.

De hidtidige diskussioner af disse krav har især gået på de behov for statistik, som hidtil har været imødekommet ved folke- og bolig-tællinger, jf. fx (11).

Der har været udbredt enighed om, at det navnlig er af betydning for den samfundsbelysende statistik, at der i det administrative system findes følgende registreringer:

- a) Et personregistersystem med en bred vifte af oplysninger: Demografiske, beskæftigelsesmæssige, sociale, uddannelsesmæssige mv.
- b) Et boligregistersystem, som også omfatter de lokaliteter i bygningsmassen, der benyttes til erhvervsformål.
- c) Et erhvervsregistersystem, som omfatter både de juridiske enheder og de lokale enheder, arbejdsstederne.

Betegnelsen registersystem skal her opfattes meget bredt, idet fx personregistersystemet kan bestå af en række registre, der føres for forskellige myndigheder. For at hvert af disse registersystemer kan betragtes som en helhed er det nødvendigt, at der findes et fælles identifikationssystem for hver af de tre typer af objekter, hvilket også er forudsætningen for, at rekombination af registeroplysninger fra forskellige myndigheder kan finde sted. Det er endvidere af afgørende betydning, at der kan etableres forbindelser mellem de tre registersystemer og deres identifikationssystemer indbyrdes:

- d) En nøgle mellem personer og de boliger, de bebor.

- e) En nøgle mellem personer og de virksomheder (arbejdsstederne), hvor de er beskæftiget.
- f) En nøgle mellem arbejdsstederne og de lokaliteter, de optager.

Er disse indholds krav opfyldt helt eller tilnærmelsesvis, har man redskaberne til at opbygge et registerstatistisk system, som på grundlag af de administrative data kan give opgørelser af folke- og bolig-tællingstype, og som tillige kan ligge til grund for væsentlige dele af den øvrige statistik vedrørende personer, boliger og lokale erhvervsenheders forhold.

5.2. Registret som statistisk model

De krav, som den statistiske anvendelse stiller til registersystemerne, falder i nogen grad sammen med kvalitetskrav, som også må være opfyldt, for at registrene kan tjene deres primære formål som hjælpemidler i den administrative proces. Hvis statistikkens krav er mere vidtgående, kan der opstå problemer; det er vigtigt at søge afklaret, i hvilket omfang sådanne problemer kan antages at gøre sig gældende.

Det første krav, der må stilles, er at der skal findes en præcis definition, der afgrænser det enkelte objekt, dvs. den enhed, man i en given sammenhæng er interesseret i at indsamle information om. Objekter kan være fysiske genstande eller personer, og de kan være mere abstrakt definerede. Der kan hensigtsmæssigt sondres imellem:

- a) Fysiske objekter
 - a1) Fysiske objekter, hvis grænser overfor andre er naturligt fastlagt; fx personer, biler, skibe
 - a2) Fysiske objekter, hvis grænser ikke er naturligt fastlagt, men bestemt af vedtagne definitioner, regler, forvaltningsakter eller sædvaner; fx ejendomme.
- b) Abstrakte objekter; fx virksomheder.
- c) Kombinationsobjekter, dvs. objekter der er sammensat af andre objekter efter givne regler; fx familier bestående af flere personer, firmaer bestående af flere virksomheder (filialer).

Det er klart, at definitionernes entydighed lettest kan opnås for fysiske og velafgrænsede objekter som personer, medens man kan få meget store vanskeligheder, når det drejer sig om abstrakte objekter. Definitionen må bl.a. tage stilling til, hvorvidt ændringer gennem tiden påvirker objekternes identitet. Der må med andre ord tages stilling til spørgsmålet om, hvornår et objekt dør og et nyt fødes: Hvis et fir-

ma får ny ejer, flytter eller skifter branche, er der da tale om et helt nyt objekt?

Man kan ikke uden videre gå ud fra, at identitetsspørgsmålene er tilfredsstillende løst i registre, som er indrettet til at løse én bestemt administrativ opgave. De i afsnit 3.3. omtalte fællesregler for erhvervsregistrering har netop til formål at løse disse problemer.

To andre vigtige krav er, at registrene må indeholde præcise og entydige identifikationsoplysninger, og at gruppen af objekter, der er beskrevet i et register, skal være afgrænset af en regel.

Når et registersystem skal benyttes som grundlag for statistiske opgørelser, betyder det at registret opfattes som en model af den gruppe objekter, som registret skal beskrive.

I mange tilfælde må en effektiv udnyttelse yderligere forudsætte, at de dataændringer, der indføres i registret, kan opfattes som model af de ændringer, objekterne undergår som følge af forskellige arter af hændelser (det forudsættes her, at der er tale om et "egentligt register", således at ajourføring af data i det hele taget foretages). Det gælder eksempelvis i befolkningsstatistikken, hvor opgørelser af hændelser i befolkningen skal kunne udledes af CPR's dataændringer, og i byggestatistikken, hvor nybyggeriets igangsættelse og færdiggørelse skal fremgå af transaktioner til bygnings- og boligregistret.

Registrene må derfor være således indrettede, at opgørelser af registrets bestand og af bevægelserne heri tidsmæssigt kan afgrænses i overensstemmelse med de tidspunkter, hvor hændelserne faktisk har fundet sted. Det er her af afgørende betydning for statistikken, at registrene kan tilvejebringe pålidelige tidsreferencer, ikke blot vedrørende tidspunktet for indførelse af en oplysning i registret (ajourføringsdato), men også om tidspunktet for den hændelse, der forårsager dataændringen. Der er imidlertid eksempler på, at de administrative registerførere ikke anser hændelsesdatoer for særligt væsentlige.

Dertil må det af hensyn til bevægelsesopgørelser kræves, at man ved registrets ajourføring kan adskille virkelige hændelser fra fejlrrettelser og mere teknisk betonedede ændringer i registeroplysningerne. Indholdet af dette sidste krav kan bedst belyses gennem nogle problemer, som rent faktisk har gjort sig gældende ved befolkningsstatistikens anvendelse af CPR, - selvom netop CPR må siges at være det registersystem, der på nuværende tidspunkt bedst imødekommer statistikproduktionens behov.

Opgørelserne over indenlandske flytninger tager udgangspunkt i oplysninger om de personer, som i en given periode får ændret adresseangivelse i CPR. I denne gruppe indgår de personer, som faktisk er

flyttet i perioden, men der indgår tillige en række persongrupper, som ikke har interesse for statistikken, nemlig bl.a.:

- 1) Personer, der er omfattet af grænsereguleringer mellem kommunerne eller kommunesammenlægninger.
- 2) Personer, hvis adresse ændres som følge af en vejnavneændring eller en omnummerering eller omkodning af vejen.
- 3) Korrektion af forkerte adressekoder.

Navnlig adresseændringer af type 2 har haft et betydeligt omfang, som er anslået til ca. 1/4 mill. personer årligt. Selvom kommunerne efter reglerne skulle give indberetningerne af disse adresseændringer en særlig markering, er det hidtil kun sket i knap halvdelen af tilfældene, og af denne grund var det ikke i de første år muligt at fremstille en brugbar generel statistik over samtlige indenlandske flytninger, men kun over den del af flytningerne, der passerer kommunegrænser.

Det skal her bemærkes, at disse problemer stort set er løst med en teknisk omlægning af CPR, som fandt sted i 1978.

De egentlige korrektioner af forkert indberettede data frembyder imidlertid et problem for statistikken, selv om man er i stand til at kende dem fra de virkelige hændelser. Korrektioner må nemlig i mange tilfælde betragtes som rettelse til tidligere indberettede hændelser: En adressekorrektion kan være en rettelse til en indberetning om flytning. En korrektion kan også annullere en tidligere indberettet hændelse, hvilket i CPR-systemet navnlig finder sted, når en person, efter at have meldt udvandring til folkeregistret, ombestemmer sig og bliver i landet.

Korrektioner til hændelser er vanskelige at behandle, fordi de i realiteten kan blive indberettet med vilkårlig lang forsinkelse. I befolkningsstatistikken har man valgt at holde dem helt uden for opgørelserne af bevægelserne, men denne fremgangsmåde er kun brugbar, fordi korrektionernes antal er ringe, sammenlignet med hændelsernes antal.

5.3. Statistikproduktion og registerorganisation

Det ligger i sagens natur, at vilkår og behov for samkøring af registerdata er stærkt afhængig af registrenes organisatoriske placering og opdeling. Derfor må statistikproducenten interessere sig for og søge at få indflydelse på den samlede registerstruktur. For Danmarks Statistiks vedkommende findes der hjemmel for en sådan indflydelse i Lov om Danmarks Statistik, §1, stk. 1, pkt. 3, og Danmarks Statistik

medvirker da også i forhandlingerne om disse spørgsmål, både ved to-sidede kontakter med andre myndigheder og gennem de såkaldte edb-samordningsorganer, der er opbygget i slutningen af 1970'erne under finansministerens ledelse og med deltagelse af statslige og kommunale myndigheder.

5.4. Forskellige typer registerdata

I det følgende vil det blive illustreret, hvordan registerdata kan inddeles i nogle kategorier efter det formål, det enkelte datum tjener i registret. Det har nemlig vist sig, at nogle af de krav, som den statistiske anvendelse stiller til registrene, knytter sig til bestemte typer af registerdata. Desuden må visse datatyper betegnes som særligt problematiske i relation til statistikken.

Et bestemt datum kan optræde i forskellige roller i forskellige registre; fx er boligens adressekode identifikation i bygnings- og boligregistret, mens den i CPR er et basisdatum.

5.4.1. Identifikationsdata

En grundlæggende forudsætning for effektiv registerdrift er som omtalt, at identifikationssystemet er entydigt. Dette forudsætter igen, at to betingelser er opfyldt, nemlig (a) at objekterne er entydigt defineret, og (b) at identifikationssystemet er præcist og hensigtsmæssigt udformet samt "robust" i forhold til almindeligt forekommende fejlmuligheder.

Hvad det første krav angår, er mulighederne for at imødekomme det stærkt varierende fra objekttype til objekttype. I nogle tilfælde (personer, biler, skibe, fly) er der efter sagens natur intet problem, i andre tilfælde (ejendomme, erhvervsenheder, bygninger osv.) må der søges opstillet operationelle definitioner eller retsregler, der på den ene side tager udgangspunkt i objektivt konstaterbare forhold, på den anden side bedst muligt modsvarer de til registreringen knyttede formål.

Kravet til identifikationssystemets udformning er af mere teknisk natur, og de styringsprocesser, som skal sikre mod dobbeltnummerering og mod fejl i anvendte numre (checkcifre) er velkendte. Spørgsmålet om, hvorvidt identifikationen skal være informationsløs eller informationsbærende, har også været debatteret i forbindelse med indretningen af de fleste nummersystemer. Det her i landet vigtigste identifikationsnummer - personnummeret - indeholder som bekendt visse informationer (fødselsdato og køn), men da disse er (næsten) uforanderlige, volder dette ikke styringsmæssige problemer. Tværtimod kan man sige,

at netop et fødselstidspunkt er en hensigtsmæssig oplysning at lade indgå i identifikationen, bl.a. for at gøre det let at huske, men det kan desværre ikke praktiseres for alle typer af objekter.

Det samlede registervæsen bygger på nogle få generelle identifikationssystemer: Personnummer, adressekode (for boliger mv.), matrikelnr. for ejendomme, virksomheds- eller firmanummer. Dette er afgørende for de faktiske integrationsmuligheder. Hver af disse identifikationssystemer må af hensyn til en effektiv drift styres ét sted, og de registre, som varetager denne funktion, får derfor en særlig betydning.

Disse styrende, primære identifikationer ligger til grund for de enkelte registers drift og for samspillet imellem disse, men de dækker ikke alle identifikationsbehov. Der anvendes fx en række udadvendte identifikationer, hvoraf navn og adresse for personer selvsagt er vigtige. Oplysning om fødselsregistreringssted er et andet vigtigt element i en personidentifikation, der bl.a. etablerer forbindelsen mellem folkeregistersystemet (inkl. CPR) og "civilstandsregistrene" (kirkebøgerne).

Identifikationerne tjener også til at udtrykke objektrelationer. For eksempel registreres i CPR relationen "person A er gift med person B" ved, at B's personnummer er anbragt i et særligt felt i person A's CPR-record (og omvendt).

En registreret objektrelation kan samtidig være af betydning som sekundær identifikation, som tilfældet fx er, når ejeren til et firma registreres i dettes record i erhvervsregistrene.

5.4.2. Tidsdata

Tiden er en afgørende dimension i vor daglige opfattelse og tolkning af fænomener i verden omkring os.

Da det statistiske register skal opbygges som en model til beskrivelse af en større eller mindre del af denne virkelighed, må tiden nødvendigvis indgå som en vigtig del af registret.

I det følgende skal der peges på forskellige former for tidsreferencer, som det kan være hensigtsmæssigt at lade indgå i et statistisk register, idet der drages en hovedsondring mellem dateringer, der tidsfæster fænomener i virkelighedens verden og dateringer, der tidsfæster vor måling eller registrering af disse.

Med hensyn til dateringer, der tidsfæster forhold af den førstnævnte art, kan man sondre mellem dateringer knyttet til selve objektets eksistens, og dateringer, der knytter sig til bestemte variable.

Dateringer, som knytter sig til objektets eksistens, er de samme,

som afgrænser objektets levetid, dvs. fødsels- og dødstidspunkt. Disse dateringer vil typisk være centrale i basisregistre, men kan også være vigtige i andre statistiske registre, hvis disse dækker så lang en periode, at der sker en tilgang og afgang af objekter.

Dateringer, som knytter sig til andre variable, må opdeles efter disse variables karakter.

Det er her vigtigt at sondre mellem variable, hvis værdi kun kan fastlægges i en periode (strømvariable) og variable, hvis værdi kan fastlægges på et tidspunkt (beholdningsvariable).

Som eksempel på strømvariable kan nævnes en persons bruttoindkomst, og det må i et statistisk register, der indeholder strømvariable, være vigtigt at datere den periode, variabelværdien referer til.

Som eksempel på beholdningsvariable kan nævnes civilstand eller uddannelse, og man kan her sondre mellem to forskellige former for datering af disse.

Dels kan man tale om en datering i diskret tid, hvor man kun kender de pågældende variables værdi på et (eller flere) givne tidspunkt(er), dels kan man tale om en datering i kontinuert tid, hvor man kender den periode, inden for hvilken de pågældende variabelværdier gælder. Ved datering i kontinuert tid får man samtidig en datering af variabelværdiændringer (hændelser), som vil være afgørende for, om der kan udarbejdes en egentlig forløbsstatistik.

Endelig kan det som nævnt være vigtigt at datere, hvornår målingen eller registreringen af et bestemt fænomen er foretaget, og der går her en hovedsondring mellem retrospektiv og simultan måling.

Ved en retrospektiv måling foretages målingen en kortere eller længere tid efter fænomenets forekomst, og måske foretages der flere retrospektive målinger af det samme forhold. I disse tilfælde er det vigtigt at kunne holde resultatet af de forskellige målinger ude fra hinanden.

Ved en simultan måling vil der normalt ikke være det samme behov for en eksplicit datering af målingen, idet denne med god tilnærmelse vil falde sammen med de dateringer, der knytter sig til de forskellige variable i registret.

5.4.3. Basisdata

Ved registermæssige basisdata forstås i det følgende sådanne registeroplysninger, som er beregnet til en mangesidet anvendelse, herunder til fremtidige, måske endnu ukendte formål. Begrebet svarer på dataplanet til det tidligere indførte begreb, basisregister.

Det mest udprægede eksempel på sådanne oplysninger er hovedparten

af CPR's dataindhold og særligt den del, som er "folkeregisterdata", fx bopæl, civilstand, nationalitet.

Udsondringen af disse gennemgående data fra øvrige registerdata har betydning i en række henseender:

- a) Lovgivningen vil hyppigt benytte basisdata som parametre (tildelingskriterier mv.) men det anses som regel ikke for nødvendigt at fastlægge det begrebsmæssige indhold gennem lovgivningen, fordi der er tale om almindeligt anerkendte begreber.
- b) Datasikkerhedsmæssigt set har de en særstilling, eftersom de efter sagens natur må få en betydelig udbredelse. Hertil kommer at de på grund af deres almene karakter som regel ikke objektivi set er særligt følsomme (jf. at enhver kan få oplysning om en række af erhvervsregistrets data.
- c) Oplysningerne må begrebsmæssigt set være "objektive" og "neutrale" og mindst muligt præget af forhold på specielle administrative områder.
- d) Datakvalitet og ajourføringsgrad må være høj, idet manglende nøjagtighed på anvendelsestidspunktet kan give anledning til administrativt besvær og ubehagelige konsekvenser for de registrerede.

Hvis oplysningerne skal opbevares i flere registre, er konsekvensen, at ajourføringsoplysninger må overføres mellem registersystemerne. Hvis de derimod skal opbevares ét og kun ét sted, er konsekvensen, at andre systemer skal hente deres data her, når de skal anvendes. I begge tilfælde vil en form for samkøring være en uundgåelig konsekvens.

5.4.4. Specialdata

Ved specialdata forstås i denne sammenhæng data, som indsamles direkte med henblik på funktionelt at indgå i bestemte administrative processer eller som er et resultat af sådanne. De mest typiske eksempler herpå skal formentlig hentes i kildeskattesystemets registre (lønmottagerfradrag, lejeværdi af egen bolig) og i de sociale systemer (dagpenge, sygesikring osv.).

Selv om der i forhold til basisoplysningerne er mange grænsetilfælde, er der klare principielle skillelinier til disse i en række henseender:

- a) Det begrebsmæssige indhold af specialdata vil oftest være fastlagt gennem specielle lovregler og administrative forskrifter.
- b) Datasikkerhedsmæssigt set vil de ofte være mere "følsomme" end

basisdata, men også genstand for mindre cirkulation.

- c) Oplysningerne må nødvendigvis være defineret i forhold til formålene og underkastet de begrebsmæssige ændringer, som følger af lovændringer mv.
- d) Datakvalitet og ajourføringsgrad bestemmes ud fra de konkrete anvendelsesbehov.
- e) Registerfunktionelt må de "fødes" i den pågældende administrationsgrens regie.

5.4.5. Baggrundsdata

Foruden de ovenfor omtalte datatyper, som altid er direkte nødvendige for den administrative proces, kan myndighederne have behov for nogle supplerende oplysninger. Når disse oplysninger indsamles systematisk, benævnes de baggrundsdata. Oplysningerne edb-registreres ikke nødvendigvis, men må alligevel siges at udgøre et led i det samlede registersystem.

Hvis brugen af baggrundsoplysningerne i konkrete tilfælde er skønsmæssig og relativ sjælden, vil det ikke være nødvendigt systematisk at kontrollere og komplettere oplysningerne i tilknytning til indsamlingen, idet det kan tænkes at kræve mindre arbejde alene at rette manglerne op under sagsbehandlingen i de tilfælde, hvor oplysningerne har betydning for afgørelsen af et bestemt spørgsmål. Baggrundsdata anvendes normalt ikke uden for det forvaltningsområde, hvor de er indsamlede som direkte grundlag for administrative processer, og er ikke umiddelbart velegnede til statistiske formål. Før disse oplysninger kan bruges i statistikken, må de i det mindste underkastes særlig grundig kontrol.

Som eksempler på sådanne baggrundsdata kan nævnes de specifikationer og oplysninger på selvangivelserne, som ikke dataregistreres hos ligningsmyndighederne, og en række af oplysninger på de tidligere vurderingsskemaer.

5.4.6. Notatdata

Disse oplysninger er af endnu løsere karakter end baggrundsdata. Det drejer sig om oplysninger, der ikke systematisk indsamles, men som fx er tilvejebragt under en sagsbehandling. Det kan være hensigtsmæssigt at registrere sådanne oplysninger, hvor de har eller kan tænkes at få fremtidig betydning i relation til den pågældende sagsbehandlende enhed.

Oplysningerne kan være edb-registrerede og da under en systema-

tisk ramme, men vil oftere have karakter af noteringer på kartotekskort eller lignende. Grænsen til egentlige sagsoplysninger er naturligvis ikke skarp.

Som eksempel kan nævnes en række notatfelter i bygnings- og boligregistret BBR, hvor den enkelte kommune helt på egen hånd bestemmer, hvad der skal indføres i registret og i hvilken form, det skal ske.

Oplysninger af denne karakter vil som regel kun være meningsfyldte for den pågældende sagsbehandlende enhed.

5.4.7. Statistikdata

De under 5.4.1.-5.4.4. nævnte data vil som hovedregel være teknisk egnede og ofte også relevante til statistikformål, og det samme kan i sjældnere tilfælde gælde for baggrundsdata. Bortset fra de former for statistik, som knytter sig snævert til et givet administrationsområde, er det imidlertid ikke sikkert, at oplysningerne frembringes i de kombinationer, der er relevante for statistiske formål. Da der kan rettes op på dette ved samkøring, får de reelle muligheder for sådan samkøring generelt set stor betydning i relation til statistiske formål. Dette indebærer igen, at identifikationssystemerne, hvis effektivitet er afgørende for samkøringsmulighederne, spiller en særlig rolle for statistikproduktionen. Kombinationsmulighederne kan udstrækkes til at omfatte traditionelle statistikdata derved, at de generelle identifikationssystemer også anvendes i forbindelse med indsamling af oplysninger gennem postenquete eller interviews.

En særlig form for statistikdata er data, som er optaget i administrative registre, men udelukkende med statistiske formål. Denne form for dataindsamling er blevet kaldt integreret dataindsamling. I den administrative proces kan de statistikdata, der indsamles på denne måde, sidestilles med baggrunds- og notatdata i den forstand, at deres kvalitet og fuldstændighed ikke har direkte og afgørende betydning for administrationen. Pålideligheden af denne type oplysninger kan derfor blive forringet.

Denne situation er ikke acceptabel for statistikproduktionen, idet de data, det drejer sig om, typisk vil være at betragte som meget centrale for statistikken, - i modsat fald ville man ikke have ønsket dem indført i registrene. Problemet er imidlertid vanskeligt at tackle.

Som eksempel på et statistikdatum af den omtalte art skal nævnes stillingskoden, som efter ønske fra Danmarks Statistik er blevet indføjet i CPR og i skatteregistrene. Stillingsbetegnelsen anføres af borgerne på selvangivelsen, og de kommunale skattemyndigheder skal sørge

for, at stillingsændringer bliver registreret. Omkring 6% af skatteyderne får på denne måde registreret ny stillingsbetegnelse og -kode hvert år, men den geografiske fordeling af disse stillingsændringer har været meget skæv, og det har været åbenbart, at enkelte kommuner ikke har levet op til deres forpligtelser på dette punkt.

For at få stillingskoden bedre oplyst og for at forbedre vedligeholdelsen af den iværksatte Danmarks Statistik i 1977 en oplysningsindsats overfor kommunerne, hvor man gjorde opmærksom på oplysningernes betydning for statistikken. I samme forbindelse pegede man på, at forskellige forvaltningsgrene i den enkelte kommune også har brug for disse oplysninger i forskellige forbindelser. Samtidig tilstillede man de kommunale skattemyndigheder opgørelser over deres egne stillingsændringer sammenholdt med forskellige gennemsnitstal.

Resultatet af denne indsats har ikke været alt for opmuntrende, selvom der er indtrådt en bedring i de kommuner, som tidligere har haft den dårligste indberetningsstandard. Man har derfor benyttet andre veje til at forbedre stillingsoplysningerne, idet man har suppleret dem med en række oplysninger fra andre kilder, som kan belyse arbejdsforholdene for større eller mindre grupper af de beskæftigede.

Trods de beskrevne vanskeligheder står det klart, at integreret dataindsamling på flere områder er en forudsætning for en udbygning af registerstatistikken. De nyeste erfaringer tyder da også på, at sådan indsamling kan gennemføres med et godt resultat, hvis man har mulighed for at ofre de nødvendige ressourcer på en hurtig og effektiv fejlkontrol og -opretning.

Således har det vist sig muligt i det såkaldte arbejdspladsprojekt at lukke nogle væsentlige huller i statistiksystemet gennem begrænsede justeringer af de administrative rutiner. De registreringer, arbejdspladsprojektet skal tilvejebringe, er dels systematiske oplysninger om, hvilke lokale arbejdssteder der eksisterer, dels oplysning om, hvem der er beskæftiget på disse arbejdssteder. Oplysningerne er afgørende for mulighederne for at fremstille geografisk beskæftigelsesstatistik, pendlingsstatistik og opgørelser om de lokale erhvervsenheders forhold, og desuden vil de i personstatistikken almindeligt anvendte erhvervsklassifikationer kunne forbedres væsentligt. Projektet er tillige et vigtigt led i den registerbaserede folke- og bolig-tælling i 1981.

Oplysningerne om de eksisterende arbejdssteder indsamles i det store og hele gennem traditionelle metoder, nemlig ved henvendelse til større private firmaer på basis af erhvervsregistret, til kommuner og amtskommuner samt til statslige myndigheder.

Det er et betydeligt vanskeligere problem at konstatere forbindelsen mellem arbejdsstederne og de beskæftigede personer. Det har imidlertid vist sig, at denne oplysning for den offentlige sektors vedkommende i meget stort omfang allerede findes registreret i de statslige og kommunale edb-lønansvisningssystemer. I projektet benytter man derfor oplysninger, som hentes direkte fra disse systemer, som dog forinden må kompletteres og systematiseres.

For de private arbejdsgiveres vedkommende indhentes nøglen mellem arbejdstager og arbejdssted gennem den årlige oplysningsseddel til skattevæsenet, som i den anledning er udvidet med et felt for arbejdssted; det er naturligvis kun arbejdsgivere med flere arbejdssteder, der skal udfylde dette felt. De foreliggende oplysninger viser, at også mange af de større private arbejdsgivere allerede har disse oplysninger inde i deres lønsystemer under en eller anden form. Danmarks Statistik udnytter i det omfang, der er mulighed for det, registreringerne i den form, hvori de foreligger hos arbejdsgiveren, idet det må antages, at datakvaliteten på denne måde bliver den bedste mulige. Det betyder, at der efter indsamlingen må udføres et harmoniseringsarbejde som led i statistikproduktionen.

Danmarks Statistik overtager selv kontrollen med oplysningerne umiddelbart efter, at de er indsamlet og registreret. Det er en følge af, at skattemyndighederne ikke hidtil har haft brug for arbejdssteds-koden i det administrative arbejde.

Projektet startede i 1980 og er endnu ikke kørt ind i tilstrækkelig grad til, at den fulde lære kan udtrages, men det kan konstateres, at det er muligt i meget stor udstrækning at komme igennem med opfølgningen af fejl, både hos registermyndighederne og oplysningsgiverne.

5.5. Dokumentation af registerdata

Som det fremgår af foregående afsnit er det for statistikproduktionen vigtigt at kende kvaliteten af de enkelte registerdata.

Den dokumentation af registerindhold, som i dag kan fås om de fleste administrative registre, giver ikke umiddelbart sådanne oplysninger. Databeskrivelserne med tilhørende regler for værdisæt viser blot, hvilket indhold der burde være i de enkelte felter, uanset om man systematisk har sikret sig kvaliteten. Resultatet kan blive, at man planlægger registerbaserede statistiksystemer, som der ikke er indholdsmæssig dækning for.

Det ville være ønskeligt, om hvert datum kunne forsynes med en kvalitetsdeklaration, der gav oplysning om indberetningskanaler, an-

vendelsen af data i den administrative proces, manuelle og automatiske kontroller, rettellesprocedurer, evt. rykning for manglende oplysninger, statistik over indberetningshyppighed og -tid mm. Deklarationen måtte som minimum kunne udskille data af typerne baggrunds- og notatdata.

I det omfang det ikke er muligt umiddelbart at fremskaffe sådan dokumentation, fordi registerføreren ikke er i besiddelse af oplysningerne, må der ved planlægning af ny registerbaseret statistik kræves gennemført pilotprojekter til konstatering af datas karakter.

Kapitel 6. Samkøring af registeroplysninger

Ved samkøring¹ forstås sammenkædning af information på objektniveau. Der kan enten være tale om, at man søger at sammenkæde information om samme objekt fra flere forskellige kilder (registre), eller at man sammenkæder oplysninger om forskellige objekter, som indgår i bestemte objektrelationer².

Det er allerede omtalt, at samkøringsteknikken har stor betydning for statistikproduktionen, når denne bygger helt eller delvis på administrative registre. I dette kapitel forsøges derfor en diskussion af de metodemæssige problemer i forbindelse med samkøring. Der tages udgangspunkt i en samling eksempler på samkøringsprojekter, som er gennemført i Danmarks Statistik; beskrivelser af disse projekter findes i bilag 2. Formålet med samkøringerne har naturligvis i alle tilfælde været at tilvejebringe nye datamaterialer, der er egnede til at give en statistisk belysning af relevante problemstillinger.

Det er imidlertid hensigtsmæssigt at foretage en klassifikation af samkøringerne efter deres formål i informationsteknisk henseende; her skal opereres med tre forskellige typer af samkøring:

Objektgruppeafgrænsning: Samkøringen har til formål af afgrænse eller konstatere én eller flere objektgrupper.

Variabelkombination: Samkøringen har til formål at kombinere variable, der stammer fra de forskellige samkøringsgrundlag.

Objektsyntese: Samkøringen har til formål at danne en ny type objekter.

I det følgende skal disse korte definitioner udfyldes. Men det skal indledningsvis bemærkes, at der forekommer glidende overgange mellem de tre samkøringstyper således at forstå, at der i et givent samkøringsprojekt kan tænkes at indgå mere end ét af de nævnte informationstekniske formål. Klassifikationen må ikke desto mindre anses for frugtbar, fordi der, som eksemplerne viser, knytter sig forskellige problemer til hvert af formålene.

¹ Begrebet anvendes her i en anden betydning end i Lov om offentlige myndigheders registre, hvor samkøring er sammenbringning af registre, der er oparbejdet til forskellige formål; registerlovens begreb omfatter således ikke almindelig opdatering.

² Denne definition omfatter ikke den såkaldte statistiske matchning, hvor man sammenkæder information fra forskellige kilder på objektniveau, men uden at det anses for nødvendigt at det er data vedr. eksakt samme objekt, eller eksakt relaterede objekter, der bringes sammen; der er tilstrækkeligt, at objekterne ligner hinanden i bestemte henseender. Statistisk matchning er så vidt vides hidtil ikke anvendt i Danmark, se fx (19).

6.1. Objektgruppeafgrænsning

En samkøring af to datamaterialer, der beskriver to objektgrupper A og B af samme type, kan resultere i en afgrænsning af fire objektgrupper:

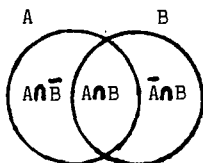
foreningsmængden af A og B ($A \cup B$)

fællesmængden af A og B ($A \cap B$)

den delmængde af A, der ikke er B ($A \cap \bar{B}$)

den delmængde af B, der ikke er A ($\bar{A} \cap B$),

jf. nedenstående figur, hvor $A \cup B$ dog ikke fremgår.



Hvis hovedformålet med samkøringen er at konstatere én eller flere af disse objektgrupper, er der tale om objektgruppeafgrænsning.

Tænker man sig eksempelvis, at A og B er den danske befolkning henholdsvis ved årets begyndelse og slutning, kan de afgrænsede grupper opfattes som:

1. Personer, der har boet i Danmark i løbet af året ($A \cup B$).
2. Personer, der har boet i Danmark hele året ($A \cap B$).
3. Personer, der er døde eller udvandret i årets løb ($A \cap \bar{B}$).
4. Personer, der er fødte eller indvandret i årets løb ($\bar{A} \cap B$).

Dette gælder dog kun med modifikationer, hvis betydning er varierende, men som under alle omstændigheder vejer tungere, jo længere det betragtede tidsrum er. For eksempel indgår personer, der både er fødte og udvandrede eller døde i året ikke i nogen af mængderne, selv om de burde indgå i nr. 1, 3 og 4.

Blandt eksemplerne i bilag 2 må navnlig samkøringerne i eksempel 12 (Analyse af dødeligheden efter erhverv) betegnes som objektgruppeafgrænsning; hovedformålet er her at konstatere, hvilke af personerne i en objektgruppe (personer i folketællingen 1970), der kan genfindes i en anden (døde 1970-75). Problemet ved denne type af samkøring er navnlig, at de stiller meget høje krav til identifikationerne og til objekternes definition, og hvis fejlriskoen er mærkbar samtidig med, at de faktiske bevægelser er små, kan resultaterne blive ganske misvisende. Ethvert tilfælde af manglende matchning på grund af forkert identifikation bliver fortolket som en hændelse eller udeblivelse af

en hændelse.

Eksempel 13 (dannelse af oplysning om udlejningsforhold i bygnings- og boligregistret) omhandler en mere problematisk samkøring af denne type, hvor man har to alvorlige fejlkilder: Identifikationssystemerne er ikke alt for sikre, og de forskellige grundoplysninger, der indgår i samkøringen, er ikke på forhånd bragt tidsmæssigt i overensstemmelse med hinanden".

Eksempler på objektgruppeafgrænsning, der har til formål at konstatere en foreningsmængde, er dannelsen af det integrerede uddannelsesregister, som beskrives i eksempel 7, og en samkøring af Kommunedatas og Københavns kommunes børnetilskudsregister, der beskrives i eksempel 1.

6.2. Variabelkombination

Ved variabelkombination forstås de tilfælde, hvor samkøringen har til formål at kombinere variable fra flere samkøringsgrundlag. Formålet kan spalttes yderligere:

1. Supplering af et datasæt med ny variabel fra et andet datasæt (horisontal integration i statistisk forstand).
2. Sammenligning af variabelværdier (fx til kontrolformål).
3. Konstatering af en ændring i en variabelværdi over en tidsperiode (vertikal integration).
4. Udvidelse eller ændring af værdisættet i en given variabel.
5. Ajourføring af registre.

Det er disse typer af samkøring, der er de mest "givende" i statistisk henseende, men samtidig dem, der skaber de største problemer.

Hvis objekterne er entydigt defineret og primæridentifikationen opfylder de for en sådan stillede krav, volder variabelkombination ingen tekniske problemer. Den mest konsekvente udnyttelse i den regelmæssige statistikproduktion er dødsfaldsstatistikken fra og med 1973 (eksempel 8).

Det er en selvfølge, at samkøring kræver fælles identifikationsvariable, men de behøver ikke nødvendigvis at være fuldstændigt identificerende. Et eksempel på anvendelse af "reserveidentifikation" med godt resultat er fødselsstatistikken fra 1973 (eksempel 9).

Mellem eksempel 8 og 9 er der naturligvis kun små praktiske nuancer. At også samkøring med langt mere ufuldstændig identifikation i betydelig udstrækning kan lade sig gøre, viser eksempel 10 (dødsfaldsstatistik 1970-1972). Ved vurdering af de overordentlig gode resulta-

ter i de tre eksempler må det tages i betragtning, at de to sæt grundoplysninger - sundhedsstyrelsens attestoplysninger og CPR-transaktioner (tidsafgrænset efter Danmarks Statistiks principper) - har en meget høj nøjagtighedsgrad, og at de to sæt oplysninger på forhånd er godt afstemt med hinanden, fordi det i stor udstrækning er de samme dokumenter, der ligger til grund for de to datasæt.

Om de enkelte typer af variabelkombination kan bemærkes:

6.2.1. Supplering af et datasæt med ny variabel fra et andet

Dette er den mest regulære form for samkøring. Problemerne vil være små, hvis det drejer sig om veldefinerede objekter og data af høj vedligeholdelsesstandard. Sådant samkøring kan substituere anden form for datafrebringelse og da med betydelige besparelser til følge (fx de nedenfor omtalte eksempler 8, 9 og 10).

Den kan også frembringe datakombinationer, som ellers er vanskelige (oplysninger om erhverv/fag kombineret med indkomst, jf. eksempel 2) eller i praksis umulige at fremskaffe, i hvert fald for større datamængder (uddannelse kombineret med indkomst, jf. eksempel 3).

Specielt må det betragtes som vigtigt, at mulighederne for at gennemføre forløbsundersøgelser, hvor bestemte persongrupper følges over længere tid (longitudinelle undersøgelser) i høj grad beror på denne type undersøgelser. Der kan her henvises til en undersøgelse af dødelighed og erhverv 1970-75 (eksempel 12, jf. (6)) og til et større forskningsprojekt, udført på Socialmedicinsk Institut ved Københavns Universitet, ang. børns opvækst under belastende vilkår, se (1).

Ved nogle statistiske analyser, som foretages på basis af samkøring, må man sige, at en tidsmæssig afstand mellem oplysningerne er en styrke. Det gælder således analysen af dødelighed efter erhverv (eksempel 12), hvor man inddrager personernes beskæftigelsessituation i 1970 som baggrundsvariabel i forbindelse med dødeligheden 1970-1975; havde man i stedet anvendt oplysninger om erhvervstilknytning på dødstidspunktet, ville det blive vanskeligt at påvise en overdødelighed i eventuelle risikofyldte erhverv, fordi udøverne af dette erhverv kunne formodes at blive pensioneret på grund af dårligt helbred inden deres død.

Et andet karakteristisk samkøringsproblem er spørgsmålet om, hvorledes man harmoniserer datasæt som nok er forskellige, men som dog indeholder overlappende informationer.

Dette problem er indgående belyst i relation til et pilotprojekt, der blev gennemført med henblik på registerfolketællingens opgørelse for 1976 (FTL-pilot 75, eksempel 4). De samme metoder er i øvrigt se-

nerne anvendt inden for mange grene af personstatistikken.

Projektets mål var at udvikle et sæt "arbejdsklassifikationer", der svarer til de traditionelle folketællingsbegreber "erhverv, fag og arbejdsstilling".

Disse tre variable er forskellige, men dog sammenbundet indbyrdes efter regler, som bl.a. medfører, at en række kombinationer er invalide.

Dette tages der i sædvanlig folketællingsmetodik højde for ved skemaets udarbejdelse (og derigennem ved dets udfyldelse), ved fortolkningen og kodningen af svarene og endelig ved en afsluttende fejlsøgning.

I den folketællingslignende undersøgelse fremkommer de tilsvarende oplysninger fra forskellige kilder, der overlapper hinanden på en sådan måde, at samme egenskab kan fremgå af forskellige grundoplysninger.

At information er overlappende betyder også, at der kan være modstrid mellem de forskellige kilder, hvis én af følgende betingelser er opfyldt:

- 1) Tidskomponenten er forskellig.
- 2) Den faktiske definition af egenskaber er forskellig.
- 3) Oplysningerne er mere eller mindre ufuldstændige og/eller upræcise.

Alle disse forhold gør sig gældende i den folketællingslignende undersøgelse. Et hovedproblem i projektet er derfor at udvikle hensigtsmæssige operationsregler, bl.a. med sigte på at gøre oplysningerne egnede til krydstabeller.

Uoverensstemmelser mellem oplysningerne kan give anledning til den opfattelse, at samkørte oplysninger er af ringere kvalitet end direkte indsamlede data, men dette er ikke nødvendigvis tilfældet. I begge tilfælde vil der i større eller mindre omfang være unøjagtigheder, i begge tilfælde må der ske en harmonisering. Forskellen er, at denne i samkøringssituationen sker samlet og på en sådan måde, at uoverensstemmelserne kommer frem i lyset, og at man har mere kontrol over harmoniseringsreglerne.

Harmoniseringsproblemet indebærer en vis sammenligning mellem data. Dette moment kommer tydeligere frem, hvis man ved en samkøring matcher oplysninger om samme eller lignende variable fra to forskellige kilder, jf. nedenfor.

6.2.2. Sammenligning af variabelværdier

Et åbenlyst formål med en samkøring kan være at kontrollere ét sæt oplysninger med et andet. Også denne problematik belyses i eksempel 4, idet den nye metode for erhvervskodning kan sammenlignes med den gamle ved samkøring af materiale fra den folketællingslignende undersøgelse med beskæftigelsestællingsoplysninger.

Denne kontrolmetode er i en vis forstand meget fintmærkende, idet den afslører enhver begrebsmæssig nuance, enhver målingsmæssig unøjagtighed, enhver tidsbestemt forskel og alle andre fejl og ufuldkomnheder. Den kan derimod ikke umiddelbart vise, hvilken af de to metoder, der er skyld i forskellene. Fortolkningen af resultaterne kan derfor være vanskelig, så meget mere, som der oftest er fejl på begge sider. Kun hvis de samlede forskelle er små, eller hvis der er mulighed for at opstille begrundede hypoteser om deres årsag, er det muligt at drage sikre konklusioner.

I det foreliggende eksempel kan det konstateres, hvilket omfang den på forhånd ventede uoverensstemmelse med deraf følgende manglende sammenlignelighed har, men i og for sig ikke hvilke oplysninger der måleteknisk er rigtigst. Dette ville kræve en supplerende undersøgelse med helt andre dataindsamlingsmetoder.

Løbende kontrol af registerstatistikken ved hjælp af supplerende dataindsamling direkte fra befolkningen og samkøring vil i stigende grad blive en nødvendighed, når større og større dele af personstatistikken bliver baseret på administrative registre. Ellers er der risiko for, at statistikken vil blive "institutionaliseret", dvs. i for høj grad belyser forholdene fra de offentlige forvaltningsorganers synsvinkel. Mere om kvalitetskontrol findes i kapitel 8.2.

6.2.3. Konstatering af ændring i variabelværdi

Formålet er her at følge udviklingen i én eller flere variable og dermed konstatere et forløb for de enkelte individer fra tidspunkt til tidspunkt (eller fra periode til periode).

Ved sådanne analyser er kravet til nøjagtighed i variabelværdierne stort, idet en hvilken som helst fejl i materialet på det ene eller det andet tidspunkt/periode vil blive opfattet på samme måde som en forskydning.

Et forsøg på at vurdere den kvantitative betydning af denne særlige fejlkilde er beskrevet i eksempel 5.

6.2.4. Udvidelse af værdisættet i en given variabel

Her tænkes bl.a. på samkøringer af datamaterialer, som for én given objektgruppe hver for sig indeholder oplysninger om én og samme variabel, men grupperet efter hver sin skala. Ved kombination af de to skalaer opstår en ny.

Der kan ikke fremlægges karakteristiske eksempler på statistiske samkøringer af denne art, selvom bl.a. det tidligere omtalte eksempel 4, FTL-pilot 75, rummer elementer heraf.

6.2.5. Ajourføring

Denne type af samkøring har til formål at indføre nye variabelværdier i stedet for forældede oplysninger. Den er velkendt fra almindelig administrativ registerdrift, men stort set tilsvarende samkøring foretages også med henblik udelukkende på dannelse af statistisk materiale. Det gælder således opgørelsen af den samlede befolkning pr. 1. januar hvert år, hvor udgangspunktet er de personer, som på dette tidspunkt står tilmeldt det centrale personregister, CPR. Denne befolkning korrigeres for eller ajourføres med meddelelser om fødsler, dødsfald, flytninger mv., som er sket før 1. januar, men som indberettes til CPR inden for de første 40 dage af det nye år; derved tager man højde for det forhold, at CPR's indhold ikke er en velegnet model af befolkningen på et givet tidspunkt, fordi hændelserne indberettes til registret med en varierende forsinkelsestid, som i gennemsnit udgør 2 uger.

6.3. Objektsyntese

Denne type samkøring har til formål at aflede oplysninger om en ny type objekter ved hjælp af data vedrørende de objekter, der indgår i det "sammensatte" objekt. Hertil kræves, at der i datamaterialerne findes variable, der angiver objektrelationerne.

Sådan samkøring vil ofte finde sted inden for materialer, der stammer fra samme kilde. I bilag 1, eksempel 6 beskrives dannelsen af objektet "vielser" ud fra objektgrupperne "viede mænd" og "viede kvinder" på grundlag af CPR-oplysninger. Et andet tilfælde er dannelse af familier på grundlag af CPR-oplysninger.

Kapitel 7: Et registerstatistisk system

7.1. Danmarks Statistiks målsætning

Spørgsmålet om, i hvilken udstrækning statistikkens krav til registersystemerne kan tænkes opfyldt, og hvor langt man derfor bør gå i retning af registerbaseret statistik, har været indgående diskuteret i mange år, ikke mindst på internationalt plan. Drøftelserne har vist en udbredt opfattelse i andre lande af, at den for statistikken fornødne relevans, kontinuitet og pålidelighed kun opnås, når de traditionelle statistiske indsamlingsmetoder benyttes.

Mange danske statistikere ser mere optimistisk på den registerbaserede statistik, og man fremhæver, at kvalitetsproblemerne med hensyn til relevans, kontinuitet og pålidelighed vedrører al slags statistik, og at statistik frembragt på traditionel måde kan være behæftet med fejl i samme omfang, der blot synes mindre fordi man har vænnet sig til dem.

I Danmark har situationen været den, at lovgivningsmagten allerede ved oprettelsen af Danmarks Statistik i 1966 har forudsat en kraftig udvikling i retning af registerbaseret statistik, især for at nedbringe borgernes og erhvervslivets indberetningsbyrde. De økonomiske forhold for institutionen har yderligere tilskyndet til en sådan udvikling. Det er ikke forkert at sige, at Danmarks Statistik siden tidligt i 1970'erne har lagt sig på registerlinien i sin strategiske planlægning. Når det har været muligt at gå denne vej, hænger det naturligvis også sammen med, at udviklingen i de administrative registre i de nordiske lande og specielt i Danmark er nået langt i retningen af at tilfredsstille statistikkens indholdsmæssige og formelle krav.

Som noget afgørende kommer hertil, at registeroplysningernes eksistens i sig selv gør indsamling af tilsvarende data særdeles problemfyldt, idet det selvsagt er vanskeligt at overbevise indberetterne om det rimelige i, at de skal afgive oplysninger til statistiske formål, som i forvejen meddeles til det offentlige i anden sammenhæng.

Registerlinien vil på nogle statistikområder indebære, at man ud fra en helhedsvurdering må vælge en anden indsamlingsmetode end den, der ellers måtte betragtes som optimal for det pågældende område, eller at man må opgive indsamling af visse oplysninger, der med traditionelle indsamlingsmetoder kunne fremskaffes til lave marginalomkostninger. Der er altså virkelig tale om et valg mellem to strategier, og valget bliver afgørende for en række andre beslutninger og for indholdet af en væsentlig del af udviklingsarbejdet.

Valget af strategi har mest klart været markeret ved beslutningen

om at basere folke- og boligtællingen 1981 udelukkende på oplysninger i administrative registre, idet man samtidig har iværksat projekter til udbygning og forbedring af dette registergrundlag.

I det følgende gives en beskrivelse af strukturen ved begyndelsen til 1980'erne i det statistiksystem, man har bygget op omkring registeroplysninger. Fremstillingen omhandler person- og familiestatistik, samt de dele af bolig- og erhvervsstatistikken, der grænser hertil. I øvrigt findes en oversigt over registergrundlaget for den egentlige erhvervsstatistik i bilag 9.

7.2. Datagrundlaget

Statistikken om personer, familier, boliger og ejendomme bygger næsten udelukkende på uddrag fra administrative registre eller andet materiale i administrationen. De vigtigste kilder er:

| | | |
|-----------------------|---|--|
| Personer og Familier: | Det centrale personregister (Indenrigsministeriet) | |
| | Skatteregistrene (Statsskattedirektoratet) | |
| | EDB-lønssystemer (Staten og kommunerne) | |
| | Sociale registre (Kommuner) | |
| | Arbejdsløshedsregister (Arbejdsdirektoratet) | |
| | Uddannelsesoplysninger (Undervisningsinstitutioner) | |
| | Attester (Sundhedsvæsnet) | |
| Boliger og ejendomme: | Bygnings- og boligregistret (Kommunerne) | |
| | Ejendomsregistret (Kommunerne) | |
| Erhvervsenheder | VIRK-registret (Toldvæsnet) | |
| | CIR-registret (Statsskattedirektoratet) | |
| | Skatteregistre (Statsskattedirektoratet) | |

De vigtigste identifikationer, der findes i disse registre, er:

- For personer (og familier): Personnr.
- For boliger (og ejendomme): Beliggenheds-adressekode
- For erhvervsenheder: Arbejdsstedskode

Der findes i et vist omfang nøgler mellem disse identifikationer i de administrative registre, og Danmarks Statistik har selv iværksat projekter til sikring af nogle af de manglende nøgler. For øjeblikket synes der kun at være uoverstigelige vanskeligheder ved at etablere forbindelsen mellem på den ene side oplysning om arbejdssteder og på den anden side oplysning om de lokaler (byggningsdele), hvor de har til huse.

Hovedreglen er, at Danmarks Statistik får stillet uddrag af registre til rådighed i en form, hvor de enkelte objekter er identifice-

rede. I enkelte tilfælde (fx statistikken vedr. modtagere af sygedagpenge) indsamles grundoplysningerne dog i form af optællinger, fx et antal tabeller pr. kommune, som derefter summeres til landsdækkende statistik i Danmarks Statistik. Anvendelsen af denne metode er på retur, fordi den kun giver små muligheder for at imødekomme andre statistikønsker end de oprindeligt forudsete, og forbindelsen til andre grunddata kan kun skabes på aggregeret niveau. Disse statistikområder ligger udenfor det samlede system, som omtales i dette kapitel.

De oplysninger, som ikke kan indhentes fra registrene, søges indsamlet enten ved statistisk dataindsamling (interviews eller post-enquete), eller der tilrettelægges såkaldt integreret dataindsamling, hvor de ekstra statistikoplysninger indarbejdes i eksisterende, administrative dataindsamlingsprocesser. I alle tilfælde drages der så vidt muligt omsorg for, at oplysningerne indeholder de omtalte identifikationer.

7.3. Systemets elementer

Systemet er opbygget på den måde, at grundoplysningerne bearbejdes og indordnes i en række statistikregistre. Dataindholdet i hvert statistikregister er defineret som de oplysninger, der er nødvendige for et bestemt statistikhovedområde, fx befolkningsstatistikken.

Der er altså tale om en formålsbestemt definition af registerbegrebet, mens det i princippet ikke er afgørende for afgrænsningen, hvilke kilder oplysningerne stammer fra.

En oversigt over statistikregistrene på personstatistikområdet og deres kilder ses i bilag 3, og i bilag 4 findes en kort beskrivelse af de enkelte registre. Bilag 5 indeholder en mere detaljeret redegørelse for befolkningsstatistikregistret.

De forskellige statistikregistre kan bindes sammen af identifikationssystemet, således at det ofte er muligt at imødekomme statistikønsker, der går på tværs af de traditionelle statistikområder.

En del af oplysningerne i systemet er imidlertid beregnet til brug inden for flere statistikområder. Disse oplysninger er indordnet i nogle såkaldte klassifikationsmoduler, som er systemets bærende del. De omtales nærmere i næste afsnit.

7.4. Klassifikationsmodulerne

Inden for den socio-demografiske statistik findes et vist antal baggrundsvariable, hvis anvendelse går igen i en række statistikområder. Eksempelvis indgår oplysninger om personers beskæftigelsesmæssige status i registerfolketællingerne og den årlige indkomst og formueta-

tistik, ligesom de anvendes ved ad hoc opgørelser på en række statistikområder. Et andet eksempel er datas samling inden for afgrænsede geografiske områder, en funktion, der indgår i en lang række statistikområder. Man har i Danmark valgt at etablere et systemberedskab, hvorved disse baggrundsoplysninger kan stilles til rådighed for alle statistikgrene samt for tværgående specialopgørelser. Dette systemberedskabs enkelte dele kaldes klassifikationsmoduler.

Der findes for tiden 6 klassifikationsmoduler i Danmarks Statistik:

1. Befolkning/datomodulet
2. Områdekodningsmodulet
3. Familie/husstandsmodulet
4. Uddannelsesklassifikationsmodulet (UKM)
5. Arbejdsstedsmodulet
6. Arbejdsklassifikationsmodulet (AKM)

En kort omtale af hvert modul findes i bilag 6, mens en mere indgående beskrivelse af arbejdsklassifikationsmodulet er optaget som bilag 7.

7.4.1. Karakteristik af klassifikationsmoduler

Et klassifikationsmodul kan som omtalt karakteriseres som en systemdel, der frembringer et bestemt sæt beslægtede baggrundsvariable til brug inden for flere statistikområder.

Klassifikationsmodulet kan generere data til et selvstændigt datasæt (register). Dette er fx tilfældet med AKM og UKM. I andre tilfælde indgår klassifikationsmodulets oplysninger i et af de eksisterende statistikregistre fx befolkningsstatistikregistret.

Klassifikationsmodulerne frembringer altid oplysninger, der er identificeret på det laveste enheds- (eller objekt-) niveau, dvs. person-, adresse- eller virksomhedsniveau, således at de kan overføres til grunddata inden for de forskellige statistikområder.

I klassifikationsmodulet frembringes de ønskede baggrundsvariable ved efter et bestemt regelsæt at sammenkæde alle tilgængelige og brugbare oplysninger inden for det pågældende emneområde. Regelsættet omdanner det uensartede grundmateriale til klassifikation af modulets objekter efter de begreber, som modulet dækker.

Regelsættet er udarbejdet på grundlag af nærmere undersøgelser af grundoplysningernes egnethed og kvalitet, og det opererer på de dele af grundoplysningerne, som er i bedst mulig overensstemmelse med klassifikationsformålet. I regelsættet indgår også stillingtagen til tvivlstilfælde, uoverensstemmelser i grundmaterialet og manglende oplysninger.

7.4.2. Krav til klassifikationsmoduler

En risiko, man løber ved at indordne statistikgrenene i et omfattende system, er, at systemet bliver for stift. Systemet må ikke lægge en spændetrøje ned over de forskellige personstatistikområder, så man altid er tvunget til at bruge de samme begreber og måleskalaer på alle områder, også når det ikke er begrundet ud fra statistikens emne eller formål. Forskelle i de inddelinger, der anvendes i forskellige statistikker, skal tværtimod forekomme, men der skal kunne gøres rede for forskellene og for deres begrundelse. Dette fører til, at man bør opsamle oplysningerne på "mindste-fælles-mangefold"-niveau, dvs. i praksis (da man ikke kan forudsige alle anvendelserne) den fineste specifikation, grundoplysningerne kan bære i de forskellige dimensioner. De oprindelige, mere specificerede grundoplysninger fra de administrative registre, som ikke er bæredygtige, skal på den anden side af hensyn til overskuelighed og dokumentation ikke opsamles.

Modulerne skal være indrettet på en sådan måde, at deres anvendelse inden for forskellige statistikområder giver resultater af størst mulig relevans. Det indebærer, at klassifikationerne skal kunne tilpasses den konkrete opgave. Der skal være mulighed for alternative inddelinger.

Det kan være hensigtsmæssigt at operere med én standardinddeling for de enkelte begreber, men der må samtidig være adgang til at anvende de underliggende elementaroplysninger i det omfang de er relevante og brugbare ved alternativ klassifikation.

Gennem anvendelsen af modulet bliver man i stand til at give en helt præcis beskrivelse af definitionsforskellene.

Kravet om fleksibilitet skal ikke nødvendigvis medføre, at modulet skal være indrettet på at imødekomme alle tænkelige behov inden for det område, modulet dækker. Hensynet til overskuelighed og dokumentation samt grundmaterialets beskaffenhed sætter visse grænser.

7.4.3. Fordele ved at anvende klassifikationsmoduler

Ideen med anvendelse af klassifikationsmodulerne bygger på nogle meget generelle målsætninger for den samlede statistik, nemlig:

- høj datakvalitet
- konsistens mellem statistikgrene
- omkostningsminimering
- systemets overskuelighed

Disse punkter forklares nærmere nedenfor.

Datakvalitet

De enkelte oplysninger skal være af så høj en kvalitet, som det er muligt indenfor de givne rammer, hvilket vil sige de administrative registre. Det betyder, at alle tilgængelige oplysninger, der har direkte eller indirekte relation til et givet emne, skal udsøges og benyttes enkeltvis eller kombineret på en sådan måde, at der opnås den størst mulige overensstemmelse med de begrebsdefinitioner, der tilstræbes i den statistiske anvendelse.

Især, når det drejer sig om registerdata, som primært er indrettet efter den administrative anvendelse, er det vigtigt, at de enkelte oplysninger vurderes og bruges i sammenhæng, således at netop den statistisk relevante del af oplysningerne udnyttes.

Konsistens

Hvis der på to statistikområder begrebsmæssigt er behov for de samme grupperinger, dvs. måling på samme skala (eller samme summationer), så er det vigtigt, at de fejl og skævheder, der uundgåeligt begås ved målingen, også er de samme inden for begge statistikområder: Målingerne skal være identiske. Det vil samtidig sige, at forskelle i randfordelinger udelukkende kan henføres til reelle forskelle, ikke til forskellige målefejl eller -skævheder. Endvidere at kvotienter meningsfuldt kan beregnes med tæller og nævner fra to forskellige statistikområder.

Økonomi

Etableringen af moduler kan også ses som en ressourcebesparende foranstaltning. Den meget omfattende bearbejdelse af store og ofte forskelligartede grundmaterialer skulle, uden modul, gennemføres flere gange, ligesom flere personer skulle have detailkendskab til såvel grunddata som reglerne for modulets opbygning. Der ligger derfor indlysende besparelser i anvendelsen af moduler.

Overskuelighed

Det er et krav til det samlede system, at det så vidt muligt skal være overskueligt og logisk opbygget. Etableringen af klassifikationsmodulerne har gjort det lettere at beskrive, hvilke faciliteter der findes i statistiksystemet. Det er nyttigt i forhold til statistikkens brugere, over for registertilsynet og over for de ansvarlige for de enkelte statistikområder i Danmarks Statistik.

Samtidig samles alle de grundlæggende beslutninger, der fører frem til en bestemt statistisk baggrundsvariabel, et bestemt sted, hvilket letter metode-dokumentationen for de enkelte statistikgrene.

7.5. Oprettelse og ajourføring af statistikregistre

Som nævnt er de fleste statistikregistre bygget på administrative registeroplysninger, men i mange tilfælde frembringer de administrative registre ikke data i de kombinationer, der skal benyttes til at tilfredsstille statistikbehovene inden for forskellige områder. Den formålsoverrettede registeropbygning betyder, at ét statistikregister ofte må opbygges på grundlag af flere forskellige administrative registre. Et vitalt led i oprettelses- og ajourføringsprocesserne er derfor personorienteret samkøring af grundoplysninger fra de forskellige kilder, som for øvrigt også kan omfatte traditionelt indsamlede spørgeskemaoplysninger, hvis registeroplysningerne ikke er tilstrækkeligt dækkende. Omvendt indeholder de registre, der er baseret på særligt indsamlede oplysninger, ofte baggrundsoplysninger fra de basisregistre (fx CPR), der har dannet grundlag for spørgeskemaernes adressering, idet det selvfølgelig tilstræbes ikke at belaste befolkningen eller den offentlige administration med indberetning af oplysninger, som i forvejen findes i Danmarks Statistik.

Statistikens behov for datakombinationer betyder, at de oplysninger, der indhentes fra ét administrativt register, ofte anvendes ved opdatering af flere statistikregistre.

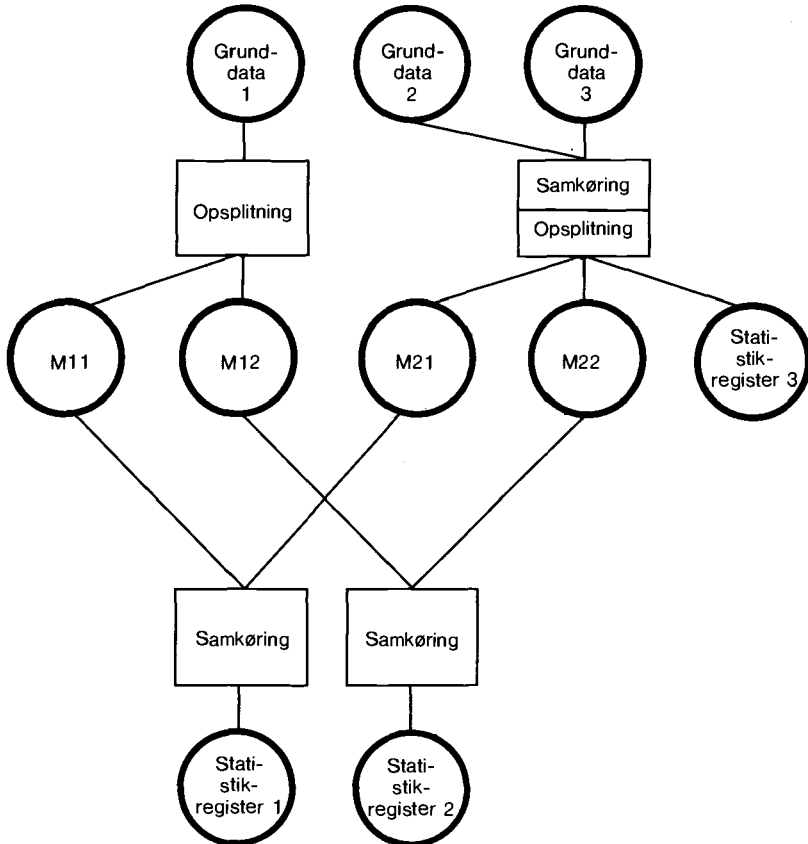
Omstående figur 1 viser, hvorledes strukturen i registerdannelsen kan tage sig ud. Man forestiller sig dannelsen af tre statistikregistre på grundlag af tre grundmaterialer. Statistikregister 1 og 2 skal indeholde oplysninger fra alle tre grunddata-sæt, mens statistikregister 3 kun skal have oplysninger fra nr. 2 og 3.

I virkeligheden er mønstret naturligvis langt mere kompliceret. Der foregår bl.a. i stort omfang overførsler fra et statistikregister til andre. En model af ajourføringsmønstret ses i figur 2. For at bevare overskueligheden er kun de vigtigste registre og dataveje medtaget i figuren. I figurens øverste bælte ses de grundlæggende administrative registre:

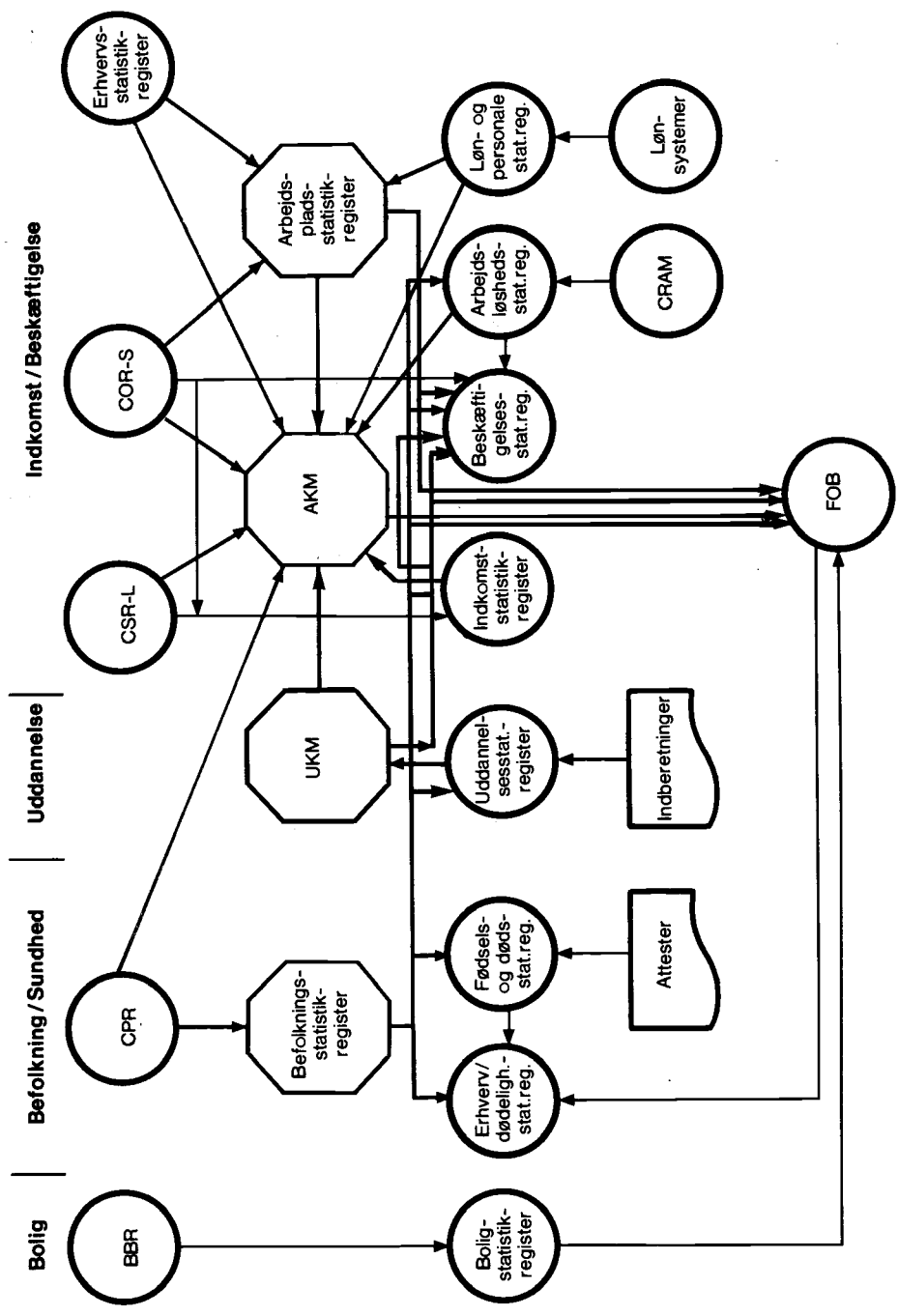
- BBR (bygnings- og boligregistret)
- CPR (det centrale personregister)
- CSR-L (det centrale skatteyderregister-ligningsdelen)
- COR-S (det centrale oplysningsseddelregister-skatteyderdelen)
- Erhvervsstatistisk registersystem.

Næste bælte er klassifikationsmodulerne, som er markeret med hver sin farve. Tredie bælte er de øvrige statistikregistre, hver for sig rettet mod et statistikområde. Under dem vises nogle mere specielle kilder, og nederst "samleregistret": folke- og bolig-tællingsregistret.

Figur 1. Fra grunddata til registre - et tænkt eksempel



Figur 2. Personstatistikken - en model af opdateringen



Erfaringerne har vist, at det er nødvendigt at lade oplysningerne fra selv det mest pålidelige administrative register underkaste en statistisk bearbejdning, som fører til en ændring af nogle oplysninger på individplanet. Denne bearbejdning kan kort siges at være begrundet i et eller flere af følgende forhold:

1. Det vil oftest være nødvendigt at korrigere registrenes tidsmæssige reference, fordi ajourføringen af registrene sker med en variende og ofte ikke ubetydelig forsinkelse.
2. Nogle registerdata har ikke en tilstrækkelig høj kvalitet, og de må derfor om muligt korrigeres (det gælder navnlig oplysninger, som kun benyttes af administrationen i visse konkrete tilfælde).
3. Det er nødvendigt af afstemme oplysninger, der stammer fra flere forskellige administrative registre.

Planlægningen af den statistiske behandling omfatter en undersøgelse af og stillingtagen til, hvilke data der har en antagelig kvalitet, og hvilke data der på den anden side er så mangelfulde, at de ikke kan anvendes til statistisk belysning af faktiske forhold.

7.6. Strukturen i det samlede statistikregistersystem

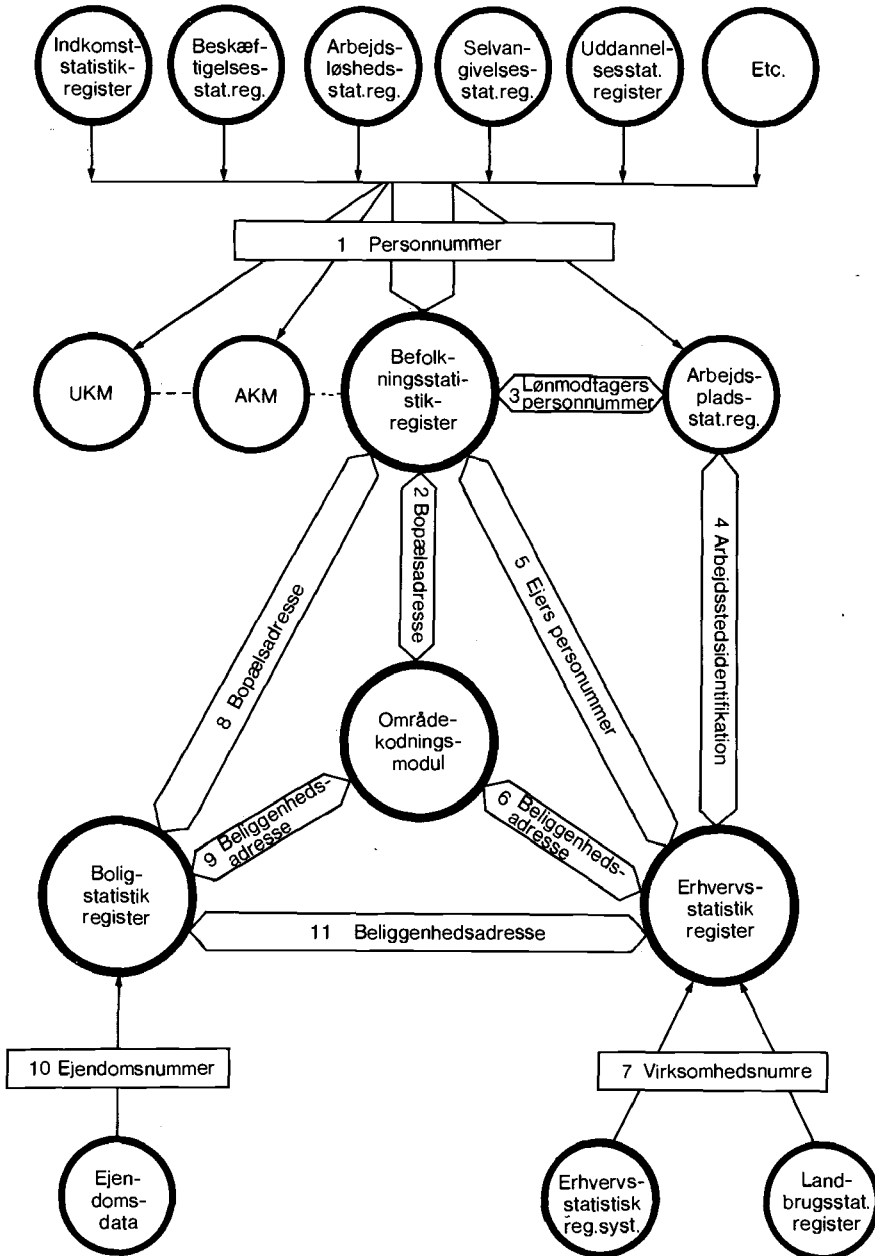
Det samlede statistikregistersystem består, som det er fremgået, af en række selvstændige registre med et dataindhold, som for hvert register sigter på et bestemt, nøje afgrænset primærformål. Men disse registre er bundet sammen til et system af nogle få identifikations-systemer. Dette forhold skaber en konsistens mellem de forskellige statistikområder, men det muliggør tillige tværgående opgørelser.

Dette afsnit giver i meget skematisk form et overblik over sammenhænge i systemet. Det er klart, at de i afsnit 7.4. omtalte "moduler" i en sådan fremstilling vil få en fremtrædende plads, da de netop er indrettet til tværgående anvendelse, hvilket samtidig betyder, at de medvirker til at binde systemets enkelte dele sammen.

Beskrivelsen knytter sig til diagrammet i figur 3, der viser systemet i hovedtræk.

Den øverste del af diagrammet viser personstatistikken, hvis grundlag er en række sideordnede registre. Personnummeret er nøgle i disse registre, og ved hjælp af dette kan man efter behov og til specialopgørelser trække på de tværgående moduler, jf. pil nr. 1. Centralt placeret blandt de tværgående moduler er befolkning/datomodulet, som er indeholdt i befolkningsstatistikregistret, og som jo giver de helt grundlæggende baggrundsplysninger om personer med bopæl i Danmark.

Fig. 3. Oversigt over registerstrukturen



Med pil nr. 2 er det angivet, at alle personoplysninger ved hjælp af områdekodningsmodulet i princippet kan grupperes efter personernes bopæl i en vilkårlig områdeinddeling. Forbindelsen er her bopælsadressen udtrykt ved adressekoden.

For lønmodtagere kan forbindelsen til arbejdsstedet etableres ved personnummeret (pil nr. 3), og forbindelsen videre til erhvervsregistret og dermed til andre virksomhedsoplysninger etableres ved en arbejdsstedsidentifikation (pil nr. 4), der består af CIR-nummer og herunder arbejdsstedskode.

For selvstændige erhvervsdrivende består forbindelsen til virksomhedsoplysningerne i ejerens personnummer (pil nr. 5), idet dette nummer er optaget i erhvervsregistret.

Også erhvervsoplysningerne kan grupperes i vilkårlige områder efter arbejdsstedets beliggenhed, idet beliggenhedsadressen findes i erhvervsstatistisk register (som er en pendant til det administrative erhvervsregister) i kodeform og dermed giver mulighed for anvendelse af områdekodningsmodulet (pil nr. 6).

I tilknytning til erhvervsstatistisk register, findes en række virksomhedsorienterede oplysninger om produktionsforhold, beskæftigelsesforhold, regnskaber mv., indordnet i to registre, det erhvervsstatistiske registersystem og landbrugsstatistikregistret. Der findes en række virksomhedsidentifikationer (CIR-nr., arbejdsstedskode, VIRK-nummer, SAF-nr.), som binder disse to registre til erhvervsstatistisk register (pil nr. 7) og dermed til det samlede registerstatistiksystem.

Som det er fremgået, findes der tillige en forbindelse mellem personstatistikregistre og boligstatistikregistret (som ikke i sig selv er et personregister) i form af personens bopælsadresse (pil nr. 8), som svarer til boligens beliggenhedsadresse. Også boligoplysningerne kan fordeles på områder (pil nr. 9). Da boligstatistikregistret omfatter ejendomsregisternummeret, er der mulighed for at tilkoble ejendomsdata (grundarealer mv.) fra ejendomsregistret (pil nr. 10).

Endelig skal nævnes muligheden af en forbindelse mellem erhvervsstatistisk register og boligstatistikregistret. Der har i de seneste år fra mange sider været udtrykt ønsker om at etablere en sådan forbindelse, så man for erhvervsenheder optaget i erhvervsregistret kunne finde arealdata mv. for tilsvarende erhvervsenheder optaget i BBR. Formelt er enhederne i de to registre identificeret på samme måde, nemlig ved kode for beliggenhedsadresse (pil nr. 11). På nuværende tidspunkt er det dog ikke lykkedes at anvise metoder til at sikre, at enhederne i de to basisregistre bliver blot nogenlunde indbyrdes konsistente.

Mulighederne for at etablere forbindelsen på ikke-aggregeret niveau må derfor betragtes som usikre.

Det kan som afslutning på denne oversigt bemærkes, at man fremover hvert år har rådighed over alle de data, som tidligere indgik i de 5-årige folke- og boligtællinger. Der er dog hverken behov for eller ressourcer til en omfattende, løbende publicering. Folke- og bolig-tællingsregistret opdateres derfor kun hvert 5. år, hvorved man samler de traditionelle folke- og bolig-tællingsdata fra flere af registrene. Herunder trækkes der på alle de omtalte tværgående moduler.

7.7. Statistisk anvendelse af eksterne personoplysninger

De tværgående moduler kan også ønskes anvendt til statistisk analysearbejde udenfor Danmarks Statistik, hvor en statistikbruger selv råder over personoplysninger om et bestemt emne. Modulerne indeholder jo nogle baggrundsoplysninger, som har meget bred anvendelighed. Desuden vil en statistikbruger ofte have særlig interesse i at benytte oplysningerne netop i den udgave, som Danmarks Statistik har oparbejdet, fordi man på denne måde umiddelbart kan sammenligne tallene fra den specielle statistik med de almindeligt tilgængelige oplysninger; det kan fx betyde, at en risikogruppe kan sammenlignes med en kontrolgruppe.

Danmarks Statistik kan ikke videregive baggrundsoplysningerne i personidentificeret form, men opgørelser af denne art kan alligevel fremstilles, hvis statistikbrugeren indgiver sit grundmateriale eller den relevante del heraf til bearbejdning i Danmarks Statistik.

Resultatet af de samkøringer, der foretages, skal altid være i en form, hvor oplysningerne ikke kan henføres til enkeltpersoner.

Kapitel 8. Fremtidig metodeudvikling

De foregående kapitler beskriver især de hidtidige erfaringer med udvikling af registerbaseret statistik og forsøger at uddrage, hvad man kan lære af disse erfaringer.

Lærdommen omhandler dels statistikkens idealkrav til de administrative registre, dels metoder til at udnytte de oplysninger, der rent faktisk er mulighed for at få. Kapitel 7 viser, hvordan man i Danmarks Statistik har søgt at bygge et sammenhængende registerbaseret statistiksystem op.

Der er imidlertid mange alvorlige metodeproblemer, som man endnu dårligt nok har identificeret, endsiige taget hul på. Nogle af dem skyldes, at idealkravene til registergrundlaget langt fra altid kan ventes opfyldt. Andre hænger sammen med, at registermetoderne åbner for en rigdom af muligheder, man ikke tidligere har haft: Hvordan skal disse muligheder udnyttes, og hvordan kan man skabe et overblik over mulighederne?

I det følgende udpeges nogle af de problemområder, som i særlig grad kræver en gennemarbejdning i de kommende år.

8.1. Helt eller delvis manglende oplysninger

Ligesom i traditionel statistikindsamling vil man i registerstatistikken hyppigt støde på et bortfaldsproblem: Det er ikke muligt at få indsamlet alle relevante data om alle objekter. I den registerbaserede statistik kan sådant bortfald forekomme i mange forskellige grader. Det hænger sammen med, at en bestemt variabel - fx oplysning om arbejdsstilling - kan stykkes sammen af elementaroplysninger fra flere kilder, således som det sker i arbejdsklassifikationsmodulet, jf. bilag 7. Her vil det forekomme, at man mangler oplysning om en person fra alle kilderne eller fra en eller flere kilder, hvor der rettelig burde være oplysning (sidstnævnte situation kunne kaldes "delvis bortfald").

Bortfaldsproblemet kunne evt. klares ved imputering, hvor man efter bestemte regler genererer en manglende oplysning på grundlag af andre foreliggende oplysninger, eller bortfaldet kunne tænkes behandlet på samme måde som i udsnitsbaserede spørgeskemaundersøgelser: Hvis det ikke kan afhjælpes, må man tage hensyn til bortfaldet i opregningen, hvorved man antager, at bortfaldet har samme fordeling som hele objektgruppen eller en bestemt del af den.

Et problem ved at bruge en sådan fremgangsmåde i registerstatistikken er, at der er tale om et sammenhængende system, hvis forskel-

lige dele skal kunne samkøres. Det ville tilsyneladende kræve, at opregningen skulle finde sted i mange tempi, nemlig hver gang man havde etableret grundlaget for en bestemt specialstatistik.

Et andet problem er, at bortfaldet sjældent er totalt, som nævnt ovenfor. Fx kan man tænke sig, at man i AKM mangler oplysning om en persons stillingsangivelse, men man har oplysning om arbejdssted (herunder branche), arbejdsomfang og løn; her findes altså en delvis belysning af personens beskæftigelsessituation, som ikke bør gå tabt.

Der bør altså ske en udnyttelse af alle de oplysningsfragmenter, man har til rådighed, og samtidig skulle det gerne være muligt at tildele alle objekter værdier af en bestemt variabel indenfor én og samme skala.

Der har endnu ikke i Danmark været arbejdet meget med mulige løsninger af dette problem og konsekvensen af sådanne løsninger. Det er i høj grad påkrævet.

Forskellige muligheder for "udfyldning" af delvis manglende oplysninger synes at kunne komme på tale afhængig af manglernes karakter. Dette kan belyses ved et par eksempler.

- a. I beskæftigelsesopgørelser har man brug for at afgøre, om en person er beskæftiget på en bestemt dato (ultimo november). Denne oplysning burde findes i skattevæsnets oplysningsseddelregister for hver person, der har modtaget løn, men den er i mange tilfælde uoplyst. I disse tilfælde kan man udnytte en anden oplysning, nemlig om personens samlede beskæftigelsesomfang i løbet af året. Hvis en person har arbejdet som lønmodtager 90% af de timer, som udgør et normalt arbejdsår, er det overvejende sandsynligt, at vedkommende var beskæftiget ultimo november; det modsatte gælder en person, der kun har været beskæftiget 10%. Man kan her tænke sig, at variabelværdien "beskæftiget ultimo november" tildeles med sandsynligheder hhv. 90% og 10%.

Denne måde at tildele manglende variabelværdier har faktisk været brugt i nogle forsøgsprojekter, men der er endnu ikke foretaget undersøgelser af, hvilke fejl dette afstedkommer i forskellige situationer. Ser man på en éndimensional fordeling af personerne efter om de er beskæftiget i november eller ej, vil den begåede fejl sikkert ikke være alvorlig, men hvis dette kriterium krydses med andre kendetegn, bliver sagen mere problematisk.

- b. Ofte finder man delvis bortfald af en mere kompliceret natur.

Et eksempel herpå, ligeledes hentet fra beskæftigelsesstatistikken: Der skal foretages en stillingsgruppering af de beskæftigede personer, og den vigtigste grundoplysning hertil er en stillingsbetegnelse, som stammer fra CPR. Når denne oplysning mangler, findes der en række andre oplysninger, som med større eller mindre sandsynlighed kan indikere en bestemt stillingsgruppe: Arbejdsstedets branche, lønmodtager/ikke lønmodtager, indkomst set i forhold til beskæftigelsesomfang. Ved vurderingen af, hvilke stillinger der må betragtes som sandsynlige for den enkelte person, må man se på samvariationen mellem den registrerede stillingsbetegnelse og de andre variable hos de personer, hvor der ikke er bortfald. I sidste instans kunne man tænke sig, at man på denne måde finder frem til sandsynligheder, hvormed personen tildeles forskellige variabelværdier (landmand, gartner eller fisker etc.).

De to eksempler adskiller sig ikke principielt. De skitserede løsninger hviler ikke på et solidt teoretisk grundlag, og der er ikke opsamlet mange erfaringer om de praktiske konsekvenser. For det sidste eksempel synes det endvidere at være kompliceret at operationalisere løsningen.

Det skal dog påpeges, at metodeproblemet har meget til fælles med det, der i den internationale statistiske litteratur går under betegnelsen statistisk matchning, jf. (19). De - noget sparsomme - erfaringer, der findes her, kan måske vise sig nyttige.

8.2. Supplerende dataindsamling: Estimation og kvalitetskontrol

Det er tidligere nævnt, at man i Danmarks Statistik har truffet beslutning om at følge en registerlinie i statistikproduktionen. Men også i en sådan statistikproduktion spiller statistisk dataindsamling ved spørgeskemaer direkte fra personer, virksomheder mv. en vigtig rolle. Det skyldes ikke blot, at nogle for statistikken vigtige oplysningstyper for tiden ikke findes i de administrative registre. Der kan også være behov for at indsamle samme oplysningstyper som de, der hentes i registrene. Formålet med sådan dataindsamling kan være dels tilrettelæggelse af estimationsprocedurer, dels løbende kvalitetskontrol.

Estimationsproblemet går i korthed ud på at finde den bedste mulige oversættelse fra en række administrativt begrundede registreringer om objekterne til de variable og skalaer, der er brug for i det statistiske system. Valget mellem forskellige mulige oversættelsesprocedurer er hidtil sket overvejende spekulativt: Man udnytter viden el-

ler formodninger om kildernes troværdighed, man betragter flerdimensionale antalsfordelinger af objekterne, som viser modstrid og samspil mellem beslægtede oplysninger fra forskellige kilder. Herefter vælges regelsættene.

Manglen ved denne spekulative metode er, at konsekvenserne af alternative regelsæt ikke holdes op mod "virkeligheden", inden valget træffes. Der er behov for, at registerkilderne kan sammenholdes, objekt for objekt, med målinger af den højeste kvalitet, der i praksis kan opnås, - det vil formentlig sige grundige stikprøveundersøgelser af rimelig stor størrelse, udført med interviewer. Det er klart, at de notoriske fejlkilder ved sådanne undersøgelser også må tages i betragtning ved estimationsarbejdet.

To områder, hvor det vil være særlig værdifuldt at kunne afprøve registeroplysningerne overfor direkte indsamlede oplysninger, skal nævnes:

- Ved fastlæggelse af regler til at korrigere for bortfaldet, jf. afsnit 8.1 ovenfor.
- I nogle tilfælde henter man generelt en bestemt variabel fra en kilde, som principielt omfatter alle objekter, mens man for en delmængde kan benytte en speciel kilde med højere datakvalitet; det gælder fx stillingsoplysningerne, hvor man principielt for alle personer har adgang til den stillingsangivelse, borgeren selv har oplyst overfor skattevæsenet, mens man fra de offentlige lønsystemer modtager oplysninger af højere kvalitet, men kun for offentligt ansatte. Hvad er konsekvensen af at blande de to typer oplysninger sammen?

Den traditionelle dataindsamling bør også bruges som en løbende kontrol på registerstatistikens indhold, så man ikke risikerer alvorlige fejlslutninger, hvis begreberne i grundregistrene forskubber sig hvilket bl.a. kan være en konsekvens af lovændringer eller ændringer i administrativ praksis.

Behovet for sådanne støtteundersøgelser er i Danmark først blevet erkendt fuldt ud på et sent tidspunkt, og der foreligger ingen erfaringer af betydning for, hvordan de skal indpasses i det samlede system.

8.3. Forløbsanalyser

Et registerbaseret statistiksystem med gennemgående identifikationer for de forskellige objekter giver nye muligheder for at fremstille statistiske analyser, hvor objekterne følges gennem tiden. Det er klart, at denne undersøgelsestype stiller store krav til identifika-

tionssystemets sikkerhed og stabilitet, og det er derfor navnlig indenfor personstatistikområdet, at mulighederne er til stede.

Forløbsanalysernes styrke er, at man kan afprøve hypoteser om sammenhæng mellem påvirkninger, som objekterne udsættes for, og senere resultater heraf. Fx inden for socialmedicinen er denne mulighed af største betydning. Mulighederne er langt fra udnyttet fuldt ud i dag¹, og det må forventes, at forløbsanalyser vil tage et stort opsving i de kommende år. I den forbindelse vil en række metodeproblemer trænge sig på.

Her skal blot nævnes spørgsmålet om, hvordan statistikregistrene bør tilrettelægges for at muliggøre, at man sammenholder på den ene side oplysninger om et historisk forløb for objekterne (fx personers erhvervshistorie gennem livet) med senere mulige virkninger (fx cancerforekomst, dødsfald). Kun på uddannelsesstatistikområdet er der i dag udviklet et beredskab for den slags undersøgelser. Også analyse- og fortolkningsproblemerne i sådanne undersøgelser vil kræve en betydelig indsats.

Det kan tilføjes, at forløbsanalyserne kan være et væsentligt grundlag for fremskrivninger og prognoser, idet man gennem disse analyser kan opgøre overgangssandsynligheder mellem to situationer, fx uddannelse og erhvervsmæssig beskæftigelse, og anvende disse i beregningen af samlet skøn over fremtidige bevægelser, eksempelvis udbud af nyuddannede på arbejdsmarkedet.

8.4. Dokumentation af statistikregistre

I afsnit 5.5 omtales problemerne i forbindelse med dokumentation af registerstatistikens kilder. Imidlertid er der tilsvarende, uløste problemer mht. selve statistiksystemet.

For at brugerne af statistikken kan få fuldt udbytte af systemet, må det kræves, at de bibringes et overblik over, hvad systemet kan frembringe. Det er et problem, som vokser i betydning, efterhånden som systemet udbygges.

Det er vigtigt at få undersøgt, hvilke krav dokumentationen skal opfylde, hvad den skal bestå af, og hvordan den skal bygges op. Det kan på forhånd anses for givet, at dokumentation vil blive så omfattende og kompleks, at den næsten nødvendigvis må bygges op som en direkte terminaltilgængelig database².

¹ Se dog fx bilag 2, eksempel 12.

² Det skal nævnes, at man i SCB i Sverige i nogle år har udført et for-tjenstfuldt, men indtil videre fejlslagent, arbejde på dette område med den såkaldte variabelkatalog, jf. (15).

BILAG

| | Side |
|---|------|
| Bilag 1: Nogle infologiske grundbegreber | 61 |
| Bilag 2: 13 eksempler på samkøring | 63 |
| Eksempel 1. Samlet anvendelse af børnetilskuds- registre | 65 |
| Eksempel 2. En analyse af bruttobevægelserne i indkomsten fra 1970 til 1971 betinget af den beskæftigelsesmæssige indpla- cering | 67 |
| Eksempel 3. Uddannelses- og indkomstoplysninger sammenholdt på udsnitbasis | 71 |
| Eksempel 4. FTL-pilot 75 (dannelse af arbejds- klassifikationsregister) | 74 |
| Eksempel 5. Undersøgelse af sikkerheden i bereg- nede bruttoerhvervsforskydninger | 81 |
| Eksempel 6. Samkøring af viede til vielser | 84 |
| Eksempel 7. Uddannelsesstatistikken af 1974 | 87 |
| Eksempel 8. Statistik over døde fra 1973 | 90 |
| Eksempel 9. Statistik over levendefødte fra 1973 | 93 |
| Eksempel 10. Oplysninger om døde 1970-75 | 96 |
| Eksempel 11. Arbejdsstyrketællingen 1975 | 100 |
| Eksempel 12. Analyse af dødeligheden efter erhverv | 106 |
| Eksempel 13. Dannelse af oplysning om udlejnings- forhold i bygnings- og boligregistret | 110 |
| Bilag 3: Statistikregistrene på personstatistikområdet og deres kilder | 115 |
| Bilag 4: De enkelte statistikregistre på personstatistik- området | 117 |
| 1. Befolkningsstatistikregistret | 118 |
| 2. Medicinsk fødsels- og dødsfaldsstatistik- register | 120 |
| 3. Folke- og bolig-tællingsregistret | 122 |
| 4. Folke- og bolig-tællingsregistret for Grønland | 123 |
| 5. Erhverv/dødelighedsstatistikregistret | 124 |
| 6. Indkomststatistikregistret | 125 |
| 7. Økonomisk statistisk register for Grønland ... | 126 |
| 8. Beskæftigelsesstatistikregistret | 127 |
| 9. Beskæftigelsesundersøgelserregistret | 128 |
| 10. Ungdomsbekæftigelsesundersøgelserregistret .. | 129 |
| 11. Arbejdsstyrkeundersøgelserregistret | 130 |
| 12. Løn- og personalestatistikregistret for den offentlige sektor | 131 |

| | Side |
|---|---------|
| 13. Arbejdsklassifikationsmodulet | 132 |
| 14. Arbejdspladsstatistikregistret | 133 |
| 15. Arbejdsløshedsstatistikregistret | 134 |
| 16. Selvangivelsesstatistikregistret | 135 |
| 17. Uddannelsesstatistikregistret | 136 |
| 18. Statistikregistret vedr. langvarigt uddannedes beskæftigelsesforhold | 137 |
| 19. Uddannelsesklassifikationsmodulet | 138 |
| 20. Bistandslovsstatistikregistret | 139 |
| 21. Kriminalstatistikregistret | 141 |
| 22. Færdselsuheldsstatistikregistret | 142 |
| 23. Huslejepanelets register | 143 |
| 24. Udsnitsarkivet | 144 |
| 25. Medicinsk dødsfaldsstatistikregister for Grønland | 145 |
| 26. Omnibusundersøgelsesregistret | 146 |
| 27. Forbrugsundersøgelsesregistret | 147 |
| Bilag 5: CPR's anvendelse i befolkningsstatistikken | 149 |
| Bilag 6: Klassifikationsmoduler i Danmarks Statistik | 157 |
| Bilag 7: Arbejdsklassifikationsmodulet (AKM) | 159 |
| Bilag 8: Milepæle i registerstatistikken siden 1968 | 167 |
| Bilag 9: Det erhvervsstatistiske registersystem | 169 |
| Bilag 10: Litteraturhenvisninger | 179 |

Bilag 1: Nogle infologiske grundbegreber

Den infologiske model tilstræber at give en informationsteoretisk begrebsramme til beskrivelse af registre og den virkelighed, de omhandler. I det følgende gives kun en kort redegørelse for nogle få, centrale begreber i modellen. En mere fyldestgørende beskrivelse findes i (16) og (17).

Et objekt er det, man i en given sammenhæng er interesseret i at indsamle information om. Objekter kan være fysiske genstande eller personer, og de kan være mere abstrakt definerede. Der kan hensigtsmæssigt sondres imellem:

- a) Fysiske afgrænsede objekter; fx personer, biler, skibe.
- b) Objekter, hvis afgrænsning mod hinanden ikke er fysisk fastlagt men bestemt af vedtagne definitioner, regler, forvaltningsakter eller sædvaner; fx erhvervsenheder (enkeltmandsvirksomhed, aktieselskab, anpartselskab), ejendomme.
- c) Kombinationsobjekter, dvs. objekter der er sammensat af andre objekter efter givne regler; fx familier bestående af flere personer, firmaer bestående af flere filialer.

Medens identiteten over tiden af objekter af hovedtype a er given, er spørgsmålet om, hvornår et objekt af typen b eller c "fødes" og "dør" kompliceret, men vigtigt for registerstyringen.

En egenskab er noget som på et bestemt tidspunkt karakteriserer et objekt. Fx kan en person have egenskaben at have blå øjne eller at være 187 cm. høj.

En objektrelation er noget som karakteriserer forholdet mellem to eller flere objekter, fx:

- person A er gift med person B
- filial C er en afdeling af firma D
- bygning E ligger på ejendom F
- person A er ejer af firma D

Objektrelationer kan som det fremgår af disse eksempler beskrive forhold mellem objekter af samme type eller mellem objekter af forskellig type.

En objektgruppe er mængden af alle objekter, som har, har haft eller vil få en given egenskab, som betegnes som den genererende egenskab. En objektgruppe er således alle personer, som på noget tidspunkt har bopæl i Danmark. Ofte vil man have brug for et tidssnit af en ob-

jektgruppe, fx alle personer, som har bopæl i Danmark den 1. januar 1978.

Det samlede objektsystem, om hvilket information ønskes samlet, består af objekterne, deres egenskaber samt objektrelationerne, der forbinder objekterne.

En variabel er en mængde af egenskaber, som på ethvert tidspunkt entydigt og udtømmende klasseinddeler en given objektgruppe. Egenskaberne må så at sige tilhøre samme dimension. Eksempler på variable er øjenfarve og legemshøjde; begge disse variable er relevante for objektgruppens personer med bopæl i Danmark, og denne gruppe betegnes relevansgruppen.

De to refererede eksempler på variable illustrerer den vigtige sondring mellem kvalitative variable (øjensfarve) og kvantitative variable (højde).

En og samme variabel kan forsynes med forskellige skalaer. Således kan variablen øjenfarve angives i en skala, der indeholder værdierne blå, grøn, grå og brun, men skalaen kan i princippet inddeles vilkårligt fint. Det skal bemærkes, at skalaen på dataplanet vil blive repræsenteret af en kodeliste.

Bilag 2: 13 eksempler på samkøring

Samlingen indeholder følgende 13 eksempler:

1. Samlet anvendelse af børnetilskudsregistrene
2. En analyse af bruttobevægelserne i indkomsten fra 1970 til 1971 betinget af den beskæftigelsesmæssige indplacering
3. Uddannelses- og indkomstoplysninger sammenholdt på udsnitbasis
4. FTL-pilot 75 (dannelse af arbejdsklassifikationsregister)
5. Undersøgelse af sikkerheden i beregnede bruttoerhvervsforskydninger
6. Samkøring af viede til vielser
7. Uddannelsesstatistikken af 1974
8. Statistik over døde fra 1973
9. Statistik over levendefødte fra 1973
10. Oplysninger om døde 1970-72
11. Arbejdsstyrketællingen 1975
12. Analyse af dødeligheden efter erhverv
13. Dannelse af oplysning om udlejningsforhold i bygnings- og boligregistret

Eksempel 1: Samlet anvendelse af børnetilskudsregistrene

1. Formål

Til brug for en undersøgelse af kvinders erhvervsaktivitet set i relation til deres familieforhold (dvs. antal børn og samlivsforhold) skal man anvende et samlet børnetilskudsregister for hele landet. Børnetilskudsregistrene er imidlertid ikke organiseret på ensartet måde i hele landet, og den her omtalte samkørsel har til formål at råde bod herpå.

2. Objektgruppe

Den objektgruppe, som tilstræbes dannet ved samkørsel, er samtlige kvinder med hjemmeboende børn under 18 år pr. 1.11.74.

3. Samkøringsgrundlaget

Objektgruppen søges afgrænset ved hjælp af 2 datamaterialer:

- 3.1. Børnetilskudsregistret for Københavns kommune pr. 1.11.74.
- 3.2. Børnetilskudsregistret for hele landet undtagen København pr. 1.11.74.

De to registre er opbygget på helt forskellig måde, men de indeholder begge oplysninger om alle børn under 18 år, som har bopæl i Danmark, dvs. som er optaget i dansk folkeregister.

4. Samkøringen

De objekter, der egentlig beskrives i registrene, er de enkelte børnetilskudssager. Beskrivelsen af en udbetaling omfatter altid en relation til den eller de tilskudsberettigede, som er barnet eller børnene, og til tilskudsmodtageren, som hyppigst - men ikke altid - vil være moderen.

I registret for København identificeres hver børnetilskudssag med tilskudsmodtagerens personnummer.

I registret for resten af landet er hver børnetilskudssag identificeret med et personnummer, som normalt er tilskudsmodtagerens, men som også kan være andet, fx den tilskudsberettigedes (barnets).

Derfor kan oplysningen om én tilskudsmodtager, hvis forhold ønskes belyst i den statistiske behandling, findes spredt på forskellige identifikationer. Behandlingen før samkøringen kræver, at oplysninger om de enkelte beløbsmodtagere samles. I virkeligheden burde man også samle oplysningerne om flere beløbsmodtagere, som måtte forekomme in-

den for samme familiekjerne. Derudover må oplysningerne for København og for resten af landet redigeres på ensartet måde.

5. Vurdering

Samkøringen er alene gennemført for tilskudsmodtagere, der er født den 16. i en måned.

Ved denne samkøring behandles registre, som er tilrettelagt udelukkende med henblik på én speciel økonomisk administrationsopgave, nemlig udbetalingen af børnetilskud. Der har ikke ved registrenes planlægning overhovedet været tænkt på en statistisk udnyttelse til belysning af tilskudsmodtagernes forhold, og derfor er identifikationsoplysningerne for hele landet undtagen København blevet u hensigtsmæssige til denne anvendelse.

Kvantitativt har problemerne dog et meget begrænset omfang, idet kun $1\frac{1}{2}$ ‰ af tilskudsmodtagerne optrådte flere steder i registret. Blandt disse var et helt forsvindende antal registreret i både registret for København og registret for det øvrige land. Samkøringens resultat må vurderes som tilfredsstillende.

Eksempel 2: En analyse af bruttobevægelserne i
indkomsten fra 1970 til 1971 betinget af den be-
skæftigelsesmæssige indplacering

1. Formål

Projektet, der er gennemført på grundlag af data fra 1970-71, har til formål at belyse de indkomstmæssige konsekvenser i forbindelse med ændring i arbejdsmarkedsstatus og beskæftigelse for den enkelte person. (Tabel 1 vises som eksempel).

2. Samkøringsgrundlaget

Oplysninger om arbejdsmarkedsstatus mv. er taget fra beskæftigelsesundersøgelserne fra maj 1970 (BU-70) og maj 1971 (BU-71), medens indkomstoplysningerne er trukket fra kildeskattesystemets slutligningsregistre for årene 1970 og 71 (KS-70 og KS-71). Samkøringen er foretaget på grundlag af CPR-nr.

3. Objektgruppen

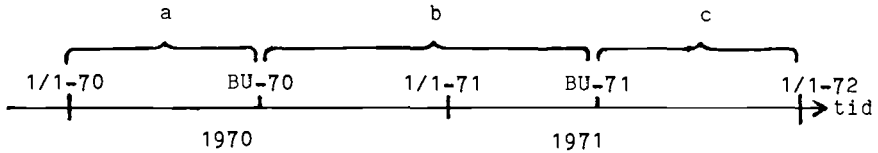
Objektgruppen er afgrænset til at være fællesmængden af de fire grundregistre, hvilket betyder at den omfatter personer, der har deltaget i beskæftigelsesundersøgelsen begge år, og til hvilke der foreligger indkomstoplysninger for såvel 1970 som 1971. Med de anvendte kriterier for dannelse af udsnit skulle objektgruppen omfatte knap 60.000 personer, men på grund af bortfald i et eller flere grundregistre udgør fællesmængden kun 47.000 personer. Bortfaldet gør sig især gældende i beskæftigelsesundersøgelserne.

4. Variablernes tidsreference

De i beskæftigelsestællingerne afgivne oplysninger refererer til en bestemt dag (tællingsdagen). Eventuelle ændringer mellem to tællinger viser sig som en forskel mellem de afgivne oplysninger, men der kan ikke opgøres mere end én ændring, og den kan i givet fald ikke tidsplaceres.

Indkomsterne refererer til et kalenderår, og en indkomstændring er forskellen mellem indkomsterne i de to år.

Problemerne omkring fortolkningen af resultaterne kan bedst belyses i følgende figur:



Erhvervsmæssig ændring kan ske på et hvilket som helst tidspunkt, men den vil påvirke resultaterne af analysen forskelligt alt efter hvor den sker. Tidsrummet for analysen er indkomstårene 1970 og 1971, og denne periode kan opdeles i tre delafsnit.

- a. perioden før BU-70
- b. perioden mellem BU-70 og BU-71
- c. perioden efter BU-71

Et erhvervsskift i periode a kan berøre indkomsten uden at der registreres en arbejdsmæssig bevægelse, fx vil et stillingsskift i april med højere indkomst til følge bevirke, at 1970-indkomsten undervurderes i relation til den beskæftigelse som registreres i BU-70. Der vil i dette tilfælde kunne observeres en overvurderet indkomststigning for personer som "tilsyneladende" ikke skifter stilling.

Stillingsskift i periode b vil blive registreret, og en evt. indkomstmæssig konsekvens vil også blive "fanget". Problemet er imidlertid, at styrken af sidstnævnte vil afhænge af, hvornår i periode b stillingsskiftet faktisk sker. Eksempelvis kan det dreje sig om en person, som er erhvervsmæssigt beskæftiget i BU-70 og pensionist i BU-71. Er ændringen sket pr. 1. januar 1971 vil der være en konsistent sammenhæng mellem arbejdsmarkedsstatus og indkomst, og man får et reelt mål for den indkomstmæssige konsekvens af beskæftigelsesændringen. Sker ændringen før 1/1-71 vil 1970-indkomsten blive et undervurderet mål for indtjeningen som erhvervsaktiv, og hvis den sker efter denne dato vil der tilsvarende ske en overvurdering af pensionistindkomsten. I begge tilfælde vil indkomstnedgangen blive undervurderet, og jo længere før eller efter 1. januar stillingsændringen sker, jo større bliver undervurderingen. Når man fremstiller statistikken, ser man ikke på enkeltindividerne, men beregner et gennemsnitligt nettoresultat for ændringen i indkomst blandt personer, som har ændret arbejdsmarkedsstatus fra aktiv til pensionist; dette gennemsnitlige nettoresultat vil imidlertid afspejle den undervurdering der sker for de enkelte personer.

Endelig vil en ændring i periode c, dvs. BU-71, påvirke indkomstforholdene uden at ændringen fremgår af BU-sammenligningen.

De nævnte forhold bliver yderligere kompliceret, når der tages hensyn til, at der kan foregå flere ændringer i beskæftigelsen mellem de to BU'er, men der vil kun blive registreret højst én bevægelse.

5. Konklusion

Som det er fremgået vil anvendelsen af indkomst pr. år føre til fejlfortolkninger, når man kæder udviklingen i dette indkomstbegreb sammen med beskæftigelsesdata for et udvalgt tidspunkt i hvert af de to år. Indkomstændringen (på årsbasis) følger i visse tilfælde kun langsomt efter en ændring i arbejdsmarkedsstatus.

I en opgørelse af indkomstudviklingen for hele objektgruppen under ét vil disse fejlfortolkninger formentlig ikke være særligt tungtvejende, men hvis man specielt analyserer den delgruppe som har ændret beskæftigelsesforhold vil de spille en betydelig større rolle.

Metoden med måling af beskæftigelsesændringer på grundlag af oplysninger fra to tidspunkter indebærer en ufuldstændig belysning af den beskæftigelsesmæssige baggrund for indkomsterhvervelsen. Der synes at være en stærk koncentration i beskæftigelsesmæssige bevægelser, sådan at en stor del af de, som skifter arbejde, gør dette mere end én gang i løbet af et år. Dette forhold gør det yderligere vanskeligt at vurdere materialets holdbarhed med hensyn til at drage konklusioner vedrørende indkomstmæssige konsekvenser af beskæftigelsesændringer.

Tabel 1. Hende bruttoindkomst samt udviklingen i denne som funktion af bruttobevægelserne mellem industrigrupper og erhverv i 1970-1971.

| | Nærings- og nydelse- industri | Tekstil- og beklæd- nings- industri | Tæ- og møbel- industri | Papir- og grafisk industri | Læder-, gummi- og kemisk industri | Sten-, ler- og glas- industri | Metal- og maskin- industri | Transport- middel- industri | Anden fremstil- lings- virksom- hed | I alt fremstil- lings- virksom- hed |
|--|---|-------------------------------------|------------------------|----------------------------|-----------------------------------|-------------------------------|----------------------------|-----------------------------|-------------------------------------|-------------------------------------|
| Udgangssituationen | 40,0 112 47,1 99 37,4 106 43,9 110 49,1 108 40,3 111 39,4 108 31,5 114 36,7 110 43,4 113 39,4 109 | | | | | | | | | |
| Afgang til: | | | | | | | | | | |
| anden erhvervsmæssig beakteligelse | 36,3 102 45,7 88 44,6 104 36,2 112 40,6 113 34,5 112 40,1 108 34,2 113 41,0 112 41,2 114 40,3 106 | | | | | | | | | |
| Heraf: | | | | | | | | | | |
| landbrug mv. | 27,5 125 15,5 123 17,4 127 - - 28,8 106 36,2 86 26,8 84 31,1 131 33,3 105 - - 26,8 109 | | | | | | | | | |
| fremstillingsvirksomhed .. | 30,9 111 35,2 99 36,2 115 39,5 107 36,6 120 34,5 109 38,4 109 34,6 106 44,6 111 45,0 123 - - | | | | | | | | | |
| Bygge- og anlægsvirks. ... | 31,1 103 31,2 109 34,7 109 39,5 122 42,3 107 41,1 107 35,9 114 27,7 114 36,8 117 38,0 112 35,8 110 | | | | | | | | | |
| handel og omsætning | 48,6 92 70,3 77 97,6 87 97,6 87 35,6 116 41,6 109 50,0 501 44,6 107 30,6 118 34,9 117 43,6 116 46,4 101 | | | | | | | | | |
| transport | 34,2 109 33,2 82 33,0 103 27,0 103 70,2 111 35,6 95 40,7 105 38,9 116 29,9 104 43,0 124 38,1 110 | | | | | | | | | |
| administration | 29,7 110 45,1 92 40,9 116 33,2 115 54,0 114 44,9 93 48,3 104 38,6 112 37,4 118 45,6 111 41,7 108 | | | | | | | | | |
| uoplyst erhverv | 17,0 95 30,1 79 15,0 112 23,5 105 30,0 20 29,8 43 18,7 76 16,0 112 23,3 70 - - 18,9 92 | | | | | | | | | |
| ikke erhvervsmæssigt be- skæftiget | 31,4 70 18,9 72 22,8 95 25,0 69 31,7 84 16,3 81 20,5 66 21,6 65 13,6 90 43,7 94 22,1 77 | | | | | | | | | |
| Forbliver i samme industri- gren | 41,8 115 50,1 104 36,1 108 46,2 111 54,2 107 42,8 111 40,6 109 31,6 115 39,7 112 43,8 113 40,9 111 | | | | | | | | | |

Anm.: Niveau angiver gennemsnitlig bruttoindkomst (kr.) i 1970

 Udvikling = gnetl. bruttoindkomst 1971 x 100
 = gnetl. bruttoindkomst 1970

Eksempel 3: Uddannelses- og indkomstoplysninger sammenholdt på udsnitbasis

1. Formål

Efter henvendelse fra undervisningsministeriet har Danmarks Statistik foretaget opgørelser af gennemsnitsindkomster for persongrupper på forskelligt uddannelsesniveau. Resultaterne skulle indgå i overvejelser vedrørende regler for tilbagebetaling af uddannelsesstøtte, og man ønskede belyst, hvorledes indkomsten for den enkelte person afhænger af uddannelsesniveau og tidsafstand fra uddannelsens afslutning.

2. Analysegrundlaget

Analysen er gennemført på udsnitbasis, således at objektgruppen er afgrænset til personer født den 15. uanset måned, dvs. ca 1/30 af hele populationen. Datagrundlaget er udtræk fra kildeskattesystemet vedrørende indkomstforhold og udtræk fra folketællingen 1970 vedrørende uddannelsesforhold. Sidstnævnte indeholder kun uddannelsesoplysninger for personer, der på tællingstidspunktet var yngre end 50 år, og de indkomst-uddannelsesmæssige sammenhænge ansås for irrelevante for de helt unge årgange, hvorfor den endelige afgrænsning af objektgruppen blev personer, der havde bopæl i Danmark ved folke- og bolig-tællingen 1970, som var 20-49 år gamle og var født den 15.

Variable i opgørelsen er bruttoindkomst, skattepligtig indkomst, uddannelseshovedgruppe, alder (substituerer periode efter afsluttet uddannelse) samt arbejdsstilling og køn. De to sidstnævnte indgår som baggrundsvariable.

Opgørelserne er foretaget på forskellige indkomstgrundlag referende til årene 1971, 1973, 1974 og 1975, men i alle tilfælde er uddannelsesoplysningerne fra 1970.

3. Effekten af det uensartede tidsreferencegrundlag

Med den foreliggende forskel i tid mellem uddannelsesoplysning og indkomstoplysning opstår der en fare for, at indkomstoplysningerne vurderes på en forkeret baggrund. I den mellemliggende periode kan uddannelser være suppleret, således at den uddannelsesmæssige baggrund for indkomsterhvervelsen er ændret. Denne risiko er naturligvis voksende jo større tidsafstand der opereres med, og den vil især gøre sig gældende i de yngste aldersklasser.

Man kan ikke direkte påvise omfanget af fejlvurderingen, og resultaterne viser da også en niveauforskell mellem de forskellige uddannel-

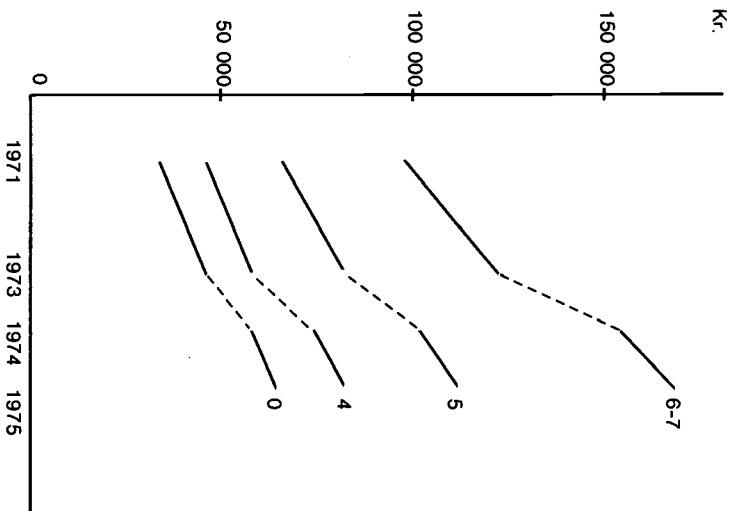
sesgrupper, jf. figur 1.

Hvis man ser på den yngste femårs aldersgruppe (fig. 2), fremgår det, at niveauforskellene er mindre, men tendensen i retning af spredning svarende til fig. 1 er der.

4. Konklusion

Opgørelserne i den foreliggende form er anvendt til belysning af sammenhængen mellem uddannelse og indkomst. Afsluttet erhvervsuddannelse må da også anses for værende uden ændringer for den enkelte person fra et vist alderstrin. En sammenligning af gennemsnitsindkomster viser nogenlunde den sammenhæng, som kunne forventes. For de yngre aldersgrupper isoleret set må det imidlertid være mere problematisk, men der kan ikke direkte påvises fejlvurderinger.

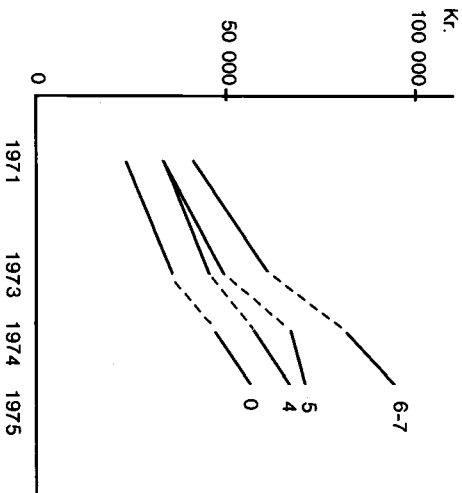
Figur 1. Gennemsnitlig bruttoindkomst for mænd i forskellige uddannelsesgrupper.



Anm. 6-7 højere uddannelse
 5 mellemuddannelse
 4 lavere uddannelse
 0 ingen eller uoplyst uddannelse

Bruttoindkomstbegrebet skifter indhold mellem 1973 og 1974, hvorfor en sammenligning over hele perioden ikke kan foretages.

Figur 2. Gennemsnitlig bruttoindkomst for mænd i alderen 20-24 år i 1970 i forskellige uddannelsesgrupper.



Anm. Se anm. til figur 1.

Eksempel 4: FTL-pilot 75 (dannelse af arbejdsklassifikationsregister)

1. Arbejdsklassifikationsregistret

Som et led i arbejdet med at foretage en registerfolketælling pr. 1. juli 1976 (FTL-76) er der gennemført et pilotprojekt (FTL-pilot¹ 75) med henblik på at afprøve mulighederne for at foretage arbejdsklassifikation af den danske befolkning på grundlag af de administrative registres dataindhold. Resultaterne af pilotundersøgelsen, der er foretaget på 1975-grundlag, skal være udgangspunkt for fastlæggelse af klassifikationsregler i den egentlige registerfolketælling 1976.

Arbejdsklassifikationsprojektets slutmål er dannelsen af et arbejdsklassifikationsregister, hvor beskæftigelsesegenskaber, udtrykt ved bestemte værdier for givne variable, for det enkelte objekt (personen) identificeres ved personnummer. Den egentlige klassifikation af FTL-objekterne sker herefter ved samkøring af klassifikationsregistret med det relevante befolkningsregister.

Selv om arbejdsklassifikationsregistrets dannelse sker i en vis sammenhæng med FTL-projektet, så skal det være generelt anvendeligt i den forstand, at det kan benyttes til lignende klassifikation af andre personorienterede identificerbare materialer.

Alle personer, der indgår i det nedenfor omtalte slutligningsregister, vil indgå i klassifikationsregistret. En yderligere objektafgrænsning foretages i forbindelse med den konkrete anvendelse af registret.

2. Grundregistre

Grundregistre i arbejdsklassifikationsprojektet var følgende:

| | |
|-------------------|-----------------------------|
| COR-S | (oplysningsseddelregistret) |
| CSR-L | (slutligningsregistret) |
| Erhvervsregistret | |
| CSR-P/CPR | (personregister) |

Objekt i COR-S er det enkelte ansættelsesforhold, pensionsforhold eller lignende (hver svarende til en oplysningsseddel). Forud for anvendelse i klassifikationssammenhæng foretages en objektsyntese (personorientering), idet oplysningsseddeldata summeres for hvert personnummer. Registrets objektgruppe er herefter alle personer, som i det

¹FTL betyder Folke Tællings Lignende opgørelse, hvilket var den oprindelige betegnelse for Registerfolketællingen 1976.

berørte kalenderår har erhvervet A-indkomst samt visse B-indkomster.

CSR-L's objektgruppe er alle personer med skattepligtig indkomst eller formue samt andre helt eller delvist skattepligtige, der er fyldt 15 år ved begyndelsen af det pågældende kalenderår.

Erhvervsregistrets objektgruppe omfatter to typer, momspligtige erhvervsenheder og indeholdelsespligtige erhvervsenheder. Registrets anvendelse baseres på udtræk af objektgruppen indeholdelsespligtige erhvervsenheder samt dannelse af indehaverregister (udtræk af de enheder i de to objektgrupper, hvor der foreligger personligt ejerforhold).

CSR-P/CPR objektgruppen er i klassifikationssammenhæng identisk med CSR-L objektgruppen.

3. Valg af variable

De traditionelle arbejdsklassifikationsvariable erhverv, fag, arbejdsstilling er centrale i beskrivelsen af den enkelte persons arbejdsmæssige egenskaber. Registergrundlaget gør det imidlertid ikke muligt at dække disse på samme måde som tidligere i forbindelse med folketællinger. Definitionerne er ændret dels i deres faktiske indhold og dels i deres referencegrundlag. Med hensyn til det sidste gælder det således, at visse grundregistres tidsreference er et helt kalenderår (selvangivelsesgrundlaget), hvilket betyder, at hovedprincippet bag arbejdsklassifikationen er væsentligste aktivitet i løbet af det berørte kalenderår. Dette princip kan imidlertid ikke fastholdes for alle variable, hvilket vil fremgå nedenfor ligesom de indholdsmæssige begrænsninger vil blive taget op.

4. Hovedvariable i arbejdsklassifikationen

Arbejdsklassifikationsregistrets dataindhold omfatter tre hovedvariable, som igen kan være grundlag for afledte variable. Disse tre variable er beskæftigelsesstatus, branche og arbejdsstillingsgruppe¹⁾.

4.1. Beskæftigelsesstatus

Med beskæftigelsesstatuskoden beskrives det enkelte objekts overordnede egenskaber i forhold til arbejdsmarkedet. Det gælder sondringen erhvervsaktiv-ude af erhverv. Endvidere omfatter værdisættet angivelse af, om erhvervsaktivitet udøves som selvstændig, medhjælpende hustru eller lønmodtager, og om personen klassificeret ude af erhverv er pensionist eller andet.

¹ Vil formentlig blive anvendt i form af egentlig arbejdsstilling og socioøkonomisk gruppe.

Grundlaget for beskæftigelsesstatuskoden er en række data fra grundregistrene COR-S (oplysningsseddelregister), CSR-L (slutligningsregister), erhvervsregistret og CSR-P/CPR (personregister). Sidstnævnte grundregistre anvendes i den forstand, at den såkaldte erhvervsstatuskode trækkes herfra. Denne kode er dannet på grundlag af data i CSR-L, således at CSR-P/CPR registrenes anvendelse i denne sammenhæng ikke er nødvendig, men blot er udtryk for den mest hensigtsmæssige fremgangsmåde.

I de anvendte grundregistre indgår en række kvalitative og kvantitative indikatorer (klassifikationsvariable) for hvilken beskæftigelsesstatuskodeværdi, der dækker den enkelte person. Disse indikatorer indgår i et regelsæt, som er grundlag for selve kodedannelsen. Detaljerne skal ikke gennemgås her, men der kan udpeges fem kvalitative hovedindikatorer, som hver især peger mod én af egenskaberne selvstændig, lønmodtager, pensionist eller andet. Disse indikatorer er følgende:

- 1a. Indehaver ifølger erhvervsregistret
- 1b. Regnskabspligt ifølge CSR-L
- 1c. Over-/underskud af virksomhed ifølge CSR-L
2. Løn ifølge COR-S
3. Offentlig pension ifølge COR-S

Hver indikator antager værdierne "ja" og "nej", og ved kombinationer fremkommer $2^5=32$ kombinationsmuligheder med varierende kompleksitet, og det kan tilføjes, at alle muligheder er fremkommet i pilotprojektet.

Indikatorvariationerne 1a, b og c i forening, 2 alene og 3 alene er éntydige i deres karakteristik af den enkelte person. Næsten på samme linie ligger kombinationer, som kun omfatter én eller to af indikatorerne 1a, b og c idet en selvstændig erhvervsdrivende ikke nødvendigvis skal have positiv indikation på de tre første indikatorer.

Mere kompliceret er kombinationer som omfatter modsatrettede indikatorer, fx indikation for selvstændig erhvervsvirksomhed samtidig med lønindkomst. Årsagen til dette kan være flere samtidige aktiviteter eller forskellige aktiviteter i forlængelse af hinanden. Her må ske en afvejning af de kvalitative indikatorer gennem anvendelse af kvantitative indikatorer som beløbsstørrelser eller beskæftigelsesperioder. I pilotprojektet er den første type valgt, idet den er mest operationel, men i relation til indkomst

ved selvstændig virksomhed er den mangelfuld, idet det skattemæssige overskudsbegreb ikke kan afvejes på lige fod med fx lønindkomst. Det indebærer, at der må ske en vis prioritering af de kvalitative indikatorer, hvilket har medført, at der lægges særlig vægt på "indehaver ifølge erhvervsregistret" og "regnskabspligt".

Resultatet af disse afvejninger bliver udtrykt i en bestemt beskæftigelsesstatuskodeværdi, som udtrykker den hovedsagelige status i løbet af et kalenderår.

Til illustration af ovennævnte indikatorkombinationer vises i tabel 1 den relative fordeling for skattepligtige personer. Med indikation for selvstændig virksomhed menes, at 1, 2 eller 3 af de tidligere omtalte typer (1a, 1b og 1c ovenfor) foreligger. Når disse tre indikatorer betragtes som én indikator er kombinationsmulighederne $2^3=8$.

Tabel 1: Relativ fordeling af skattepligtige personer efter kombinationer af klassifikationsindikatorer.

| Kombination løbenr. | Indikation for selvst. virksomhed | Lønind- komst | Off. pens- indkomst | Relativ fordeling af alle | Andel klass. som selvstændig |
|------------------------|---|------------------|------------------------|---------------------------------|------------------------------------|
| | | | | | pct. |
| 1 | + | + | + | 5,6 | 96,0 |
| 2 | + | + | + | 53,5 | 0,0 |
| 3 | + | + | + | 15,5 | 0,1 |
| 4 | + | + | + | 4,5 | 45,2 |
| 5 | + | + | + | 1,5 | 44,4 |
| 6 | + | + | + | 1,8 | - |
| 7 | + | + | + | 0,3 | 32,2 |
| 8 | + | + | + | 17,4 | 0,2 |
| | | | | 100,0 | 8,1 |

I tabellen er vist, hvor stor en andel af personerne i de enkelte kombinationsgrupper, der endeligt bliver klassificeret som selvstændige. Det kan tilføjes, at når andelen i gruppe 1 afviger en del fra 100 pct., så skyldes det især to forhold. Hvis der i erhvervsregistret optræder mere end én indehaver til samme firma skærpes kravene til andre indikatorer for at resultatet bliver klassifikation som selvstændig. Endvidere er der også visse beløbsafvejninger overskud/underskud i forhold til A-indkomst for personer som ikke forekommer i erhvervsregistret. For gruppe 3 og 8 gælder det, at visse personer klassificeres på gammelt slutligningsgrundlag fordi slutligning for det relevante år ikke er

indberettet. Der kan være flere årsager til dette, bl.a. udvandring og dødsfald i årets løb.

4.2. Branche

Brancheklassifikation er gennemført for alle personer, der ifølge beskæftigelsesstatuskoden er erhvervsaktive. Grundregistrene er i denne sammenhæng erhvervsregistret, COR-S og CSR-P/CPR. I den henseende foretages en opdeling svarende til de foreliggende muligheder.

4.2.1. Selvstændige, registreret i erhvervsregistret

For denne gruppe udtrækkes en branchekode fra erhvervsregistret. Det vil være den pr. udtrækstidspunktet gældende, hvilket betyder, at der i beskedent omfang kan være uoverensstemmelse mellem den udtrukne kode og den i året væsentligste branche som følge af brancheskift. Et andet forhold, der må tages højde for, er muligheden for, at samme person er medindehaver af flere firmaer.

4.2.2. Selvstændige, der ikke er registreret i erhvervsregistret

Disse personer kan ikke umiddelbart brancheklassificeres. Den eneste mulighed er anvendelse af stillingskoden (jf. afsnit 4.3). Denne gruppe udgjorde i pilotprojektet et ret stort antal, men efter en vis tilretning af erhvervsregisterudtrækket vil den formentlig omfatte ca. 10 pct. af de, der endeligt klassificeres som selvstændige.

4.2.3. Lønmodtagere

Fra COR-S er udtrukket det CIR-nr., som dækker den største lønudbetaling for den enkelte person. Dette CIR-nr. er grundlag for udtræk af en branchekode fra erhvervsregistret. På den måde brancheklassificeres alle lønmodtagere efter den branche, hvor den største løn er udbetalt, og lønbeløbet er således kriterium for valg mellem flere muligheder.

Med anvendelsen af erhvervsregistret som branchekodegrundlag med indgang via data, som genfindes i COR-S, vil klassifikationen følge den enhedsopdeling, der gælder ved firmaets afregning af indeholdt kildeskat. Her er hovedprincippet, at der afregnes på firmaniveau, således at der kan ske afregning for flere virksomheder (i samme eller forskellige brancher) over samme CIR-nr. Det betyder, at alle ansatte

i firmaet vil blive klassificeret under firmaets hovedbranche (største omsætning) også selv om de er beskæftiget i en virksomhed, hvis branche afviger fra hovedbranchen.

Specielt indenfor den offentlige sektor, hvor der anvendes fællesafregning i meget udstrakt grad betyder det, at der kun foretages brancheklassifikation på hovedgruppeniveau. Her udgør den offentlige sektor én gruppe. Klassifikationen af lønmodtagere har i pilotprojektet kunnet gennemføres for 97,7 pct. af lønmodtagerne.

4.3. Arbejdsstillingsgruppe

Grundlaget for klassifikation efter stillingsgruppe er stillingsbetegnelserne, som ajourføres gennem rettelser på selvangivelsen. Tidsreferencen vil således være tidspunktet for selvangivelsens udfyldelse. Denne betegnelse omsættes til en trecifret kode, og såvel betegnelse som kode vedligeholdes i CSR-P og CPR. Det ene af disse registre er således grundregister i denne sammenhæng. Stillingsgrupperne dannes med udgangspunkt i kombinationer af stillingskode og beskæftigelsesstatuskode, således at tidligere omtalte grundregistre også indirekte indgår.

Ved dannelsen af stillingsgruppekode for en given person trækkes således på en variabel dækkende den hovedsagelige beskæftigelse i løbet af et kalenderår og en variabel dækkende et tidspunkt 1-1 1/2 måned efter udløbet af det berørte år. Dette forhold vil i visse tilfælde give anledning til koblinger af inkonsistente data, som ikke refererer til samme beskæftigelse, men de vil kun i få tilfælde vise sig som umulige kombinationer. Derimod vil de give anledning til en drejning væk fra hovedprincippet om klassifikation efter væsentligste beskæftigelse i årets løb.

Tilsvarende problemer vil opstå i forbindelse med samtidig dobbeltbeskæftigelse, hvor stillingskoden kan referere til én beskæftigelse og beskæftigelsesstatuskoden til en anden.

5. Den samlede arbejdsklassifikation

De tre grundbegreber er beskæftigelsesstatus, branche og arbejdsstillingsgruppe, og de grundlæggende registerdata for dannelsen af disse er væsentligste indkomstart i forhold til samlet indkomst, erhvervsregistrets branchekode og CPR-stillingskoden. Som det er fremgået, kan klassifikation efter disse tre variable gå i forskellig retning, således at den samlede klassifikation kan være synligt eller usynligt in-

konsistent. Med synlig inkonsistens tænkes på tilfælde, hvor de forekommende værdier for de tre klassifikationsvariable er åbenbart uforenelige, medens den usynlige inkonsistens gør sig gældende, hvor den forekommende kombination af variabelværdier er mulig men ikke korrekt, fordi der ikke refereres til samme beskæftigelse.

Med valget af princippet om at branchekode efter firmahovedbranche kan der forekomme umiddelbar inkonsistent sammensætning af variabelværdier som følge af, at den enkelte persons beskæftigelse ikke er dækket gennem den foretagne klassifikation. Eksempelvis kan det dreje sig om en mølleriarbejder beskæftiget på en dampmølle som indgår i firmaet DB. Klassifikationen af en sådan person vil være mølleriarbejder i branchen detailhandel. Det vil være en umiddelbart inkonsistent kombination vurderet ud fra den karakteristisk, som gives af den pågældendes beskæftigelse, men med den anvendte definition af branche vil den ikke være inkonsistent som personkarakteristik. Det betyder imidlertid, at man må være opmærksom på, hvad der egentlig ligger i den foretagne klassifikation, nemlig at det er en mølleriarbejder, som er beskæftiget i et firma, hvis væsentligste aktivitet ligger indenfor brancheområdet detailhandel.

6. Registerbaseret arbejdsklassifikation sammenlignet med klassifikation efter traditionelle metoder

I pilotprojektet er gennemført en sammenligning af resultaterne fra den registerbaserede arbejdsklassifikation med klassifikationen foretaget i beskæftigelsesundersøgelsen. Sidstnævnte indeholder den klassifikation, som er foretaget på grundlag af den enkelte persons egne angivelser og dermed valg mellem evt. flere muligheder.

Sammenligningen afslører naturligvis uoverensstemmelser som følge af begrebsforskelle (tidsreference, firmaprincip mm.), men det er også åbenbart, at klassifikationen ikke sker efter samme principper. I registerklassifikationen indgår en række komponenter (indikatorer), og der fastlægges regler for, hvilke klassifikationer et givet sæt af indikatorer fører frem til. Ved samkøring af grundregistrene er der skabt mulighed for ensartede klassifikationsregler. Ved den traditionelle klassifikation, som vel egentlig sker på samme indikatorgrundlag, vil afvejningen af de enkelte indikatorer mod hinanden være underkastet en vis subjektivt betinget variation.

Eksempel 5: Undersøgelse af sikkerheden i beregnete bruttoerhvervsforskydninger

1. Bruttoforskydningsopgørelser

Grundlaget for opgørelse af bruttobevægelserne mellem erhvervsgrupper er oplysninger afgivet ved beskæftigelsesundersøgelserne, som Danmarks Statistik har afholdt siden 1969. Disse undersøgelser er baseret på personudsnit udvalgt på en sådan måde, at der i et vist omfang er personsammenfald mellem undersøgelser foretaget med et års mellemrum. På grundlag af CPR-nr. samkøres to sådanne undersøgelser, og man får to sæt beskæftigelsesdata for den enkelte person med reference til et givet tællingstidspunkt henholdsvis tællingstidspunktet ét år senere. Derved foretages en afgrænsning af objektgruppen til personer, der har deltaget i begge beskæftigelsesundersøgelser, og for disse foretages en vertikal variabelintegration. Ved parvis sammenligning af de enkelte data kan det undersøges om der foreligger ændringer (forskydninger) i henseende til de enkelte beskæftigelsesvariable. Resultater fra sådanne opgørelser er meddelt i Statistiske Efterretninger 1972 nr. 73 og 1975 nr. 28.

2. Undersøgelse af sikkerhed

Hovedsigtet med bruttoforskydningsopgørelserne har været at belyse bevægelserne til og fra erhvervsgrupper, og det er sket ved for den enkelte person at sammenligne erhvervs-koden ved første tælling med erhvervs-koden fra tællingen året efter. En forskel i kodeværdi vil så blive registreret som en bevægelse. Det betyder, at ændringer, der forekommer alene som følge af en uensartet fortolkning (personens egen beskrivelse, kodningen, registreringen) af et givet erhverv, vil fremstå som en bevægelse. Ved bestandsopgørelser vil fortolkningsuoverensstemmelser tendere mod at ophæve hinanden i resultaterne, medens de i opgørelser af bevægelser énsidigt trækker i retning af en overestimering.

For at belyse denne overestimering gennemførtes et projekt, hvor man foretog en samtidig kodning af erhvervsoplysninger afgivet i november 1973 og oktober 1974. Ved denne rekodning var man især opmærksom på, om erhvervet var det samme på de to tidspunkter.

Personsammenfaldet mellem beskæftigelsesundersøgelserne i november 1973 og oktober 1974 er 48.000, begrænset i forhold til de oprindelige udsnit (60.000) som følge af bortfald og aldersforskydning. Fra den oprindelige objektgruppe blev udtrukket en stikprøve på 8.150 per-

soner i rekodningsprojektet. Disse personer er kendetegnet ved at have fødselsdag på en given dag i en given måned.

3. Resultaterne af rekodningen

Blandt de udtagne personer er analysen især relevant for de, som var beskæftiget på begge tidspunkter. I tabel 1 vises resultatet for de 4.909 personer i udsnittet som opfyldte denne betingelse. Det fremgår her, at 4.105 personer på begge tidspunkter fik samme erhvervskode i forbindelse med rekodningen som ved den oprindelige kodning af hvert års oplysninger for sig, 678 personer fik ændret den ene erhvervskode, og 126 fik ændret begge koder. I en del tilfælde er der dog tale om ændringer til/fra uoplyst.

I tabel 2 fremgår rekodningseffekten på bruttoforskydningerne i opregnet form for de største erhvervshovedgrupper. Den samlede afgang fra disse fem grupper er på det korrigerede grundlag reduceret med 18 pct. i forhold til den oprindelige opgørelse. Denne bruttoafgang rettes mod andre erhvervshovedgrupper, uoplyst erhverv, værnepligt, arbejdsløshed og ude af erhverv. Omvendt er tilgangsbevægelsen til de fem grupper under ét reduceret med 15 pct.

4. Konklusion

Rekodningsprojektet har vist, at uensartet fortolkning af erhvervsoplysningen på beskæftigelsesundersøgelsesskemaet medfører en overestimering af bruttobevægelserne i forhold til erhvervshovedgruppe med ca. 20 pct. Endvidere har det afsløret en usikkerhed omkring visse bevægelser (stigende af- og tilgange i forhold til uoplyst erhverv), hvilket betyder, at den overestimering, der reelt er tale om ligger noget over de 20 pct.

Det skal understreges, at der kun er foretaget analyse på bevægelser mellem erhvervshovedgrupper.

Tabel 1. Beskæftigede i begge beskæftigelsesundersøgelserne nov. 73 og okt. 74 fordelt efter rekodningsresultatet.

| | 1974 | | I alt |
|---|---|--|-------|
| | Ny erhvervs- kode lig med gammel erhvervs- kode | Ny erhvervs- kode forskellig fra gammel erhvervs- kode | |
| 1973: | | | |
| Ny erhvervs- kode lig med gammel erhvervs- kode | 4.105 | 154 | 4.259 |
| Ny erhvervs- kode forskellig fra gammel erhvervs- kode | 524 | 126 | 650 |
| I alt | 4.629 | 280 | 4.909 |

Anm.: I tallene indgår ændringer som følge af, at den oprindelige kode er ændret til uoplyst, ændret fra uoplyst eller har været uoplyst i såvel oprindelig som i ny kode. Det drejer sig om 115 i 1973 og 108 i 1974.

Tabel 2. Samlet af- og tilgang i relation til udvalgte erhvervs-
hovedgrupper mellem beskæftigelsesundersøgelserne afholdt
i perioden november 1973 til oktober 1974, oprindelig
opgørelse og opgørelse korrigeret under hensyn til rekod-
ningsresultaterne

| | Oprindelig brutto- forskydning | Korrigeret brutto- forskydning |
|-------------------------|--------------------------------------|--------------------------------------|
| Afgang i alt | | |
| fra landbrug mv. | 28.953 | 24.900 |
| fremstillingsvirksomhed | 134.499 | 112.979 |
| bygge- og anlægsvirks. | 49.071 | 38.766 |
| handel og omsætning | 82.617 | 61.137 |
| offentlige ydelser | 62.611 | 55.724 |
| Tilgang i alt | | |
| til landbrug mv. | 27.575 | 25.369 |
| fremstillingsvirksomhed | 103.970 | 83.176 |
| bygge- og anlægsvirks. | 37.854 | 31.797 |
| handel og omsætning | 74.868 | 59.894 |
| offentlige ydelser | 91.644 | 82.480 |

Eksempel 6: Samkøring af viede til vielser

1. Vielsesstatistikken indtil 1978

Vielsesstatistikken har til formål at belyse dels "vielsestilbøjeligheden" blandt befolkningen med bopæl i Danmark og Grønland og i forskellige delmængder heraf, dels vielsernes struktur (dvs. oplysninger om vielsesmyndighed, begge parters alder mv.).

2. Objektgrupper

I vielsesstatistikken opereres med to forskellige objekttyper og dermed også to objektgrupper:

2.1 Viede personer

Objektet er her personen, og objektgruppen består af samtlige personer (mænd og kvinder) med bopæl i Danmark og Grønland, som indgår ægteskab.

2.2 Vielser

Objektet er her vielsen, og objektgruppen omfatter samtlige vielser, hvor mindst én af ægtefællerne har bopæl i Danmark eller Grønland på vielsestidspunktet.

Objekttypen vielse er sammensat af to objekter af typen viet person, idet dog den ene person kan have bopæl i udlandet og derfor ikke eksistere i objektgruppen 2.1.

3. Samkøringsgrundlaget

Oplysninger om de viede personer fås direkte fra CPR, og disse oplysninger danner tillige grundlag for opgørelserne af vielser. Samkøringsgrundlaget kan siges at bestå af:

3.1 Oplysninger om viede mænd
(med bopæl i Danmark/Grønland)

3.2 Oplysninger om viede kvinder
(med bopæl i Danmark/Grønland)

Begge materialer indeholder personnummer for den viede og henvisningsnummer, dvs. personnummer for ægtefællen.

4. Samkøringens formål

Systemet gav indtil 1978 ikke samlede oplysninger om hver vielse, idet de viede indberettes hver for sig og ikke nødvendigvis samtidig

(under det senere CPR-system, CPR-78, skal folkeregistrene derimod indberette vielser). Derfor kunne vielsesstatistikken kun frembringes ved, at oplysningerne om de viede personer kombineres. Hertil benyttes personnumrene og henvisningsnumrene som nøgler.

Samkøringen foregår i to trin, idet man først matcher par, hvor der er overensstemmelse mellem mandens personnummer og kvindens henvisningsnummer (samkøring 1). De, der ikke matches i første omgang, indgår i samkøring 2, hvor der dannes par ved overensstemmelse mellem kvindens personnummer og mandens henvisningsnummer.

Samkørslen sker for et år ad gangen.

5. Vurdering af samkøringen

I 1976 omfattede samkøringen knap 31.000 mænd og lige så mange kvinder, jf. nedenstående oversigt. I alt kunne 97,4% af de viede matches, næsten alle i den første samkøring.

| | <u>Mænd</u> | <u>Kvinder</u> |
|-----------------------|---------------|----------------|
| Antal viede i alt | 30.888 | 30.875 |
| Heraf: | | |
| Matchet i samkøring 1 | 29.975 | |
| Matchet i samkøring 2 | <u>116</u> | |
| Matchet i alt | <u>30.091</u> | |
| Ikke matchet | 797 | 784 |

Oplysningerne om de viede, som ikke kan matches, udsendes til folkeregistrene for en nærmere undersøgelse. En vigtig årsag til uoverensstemmelserne er naturligvis, at der ikke kan og ej heller skal kunne findes to ægtefæller, hvis den ene er bosat i udlandet. Det samlede antal vielser 1976 blev 31.192, og det fremkommer på følgende måde:

| Brudgom \ Brud | :"Viede Kvinder": :(dvs. bopæl Dan- mark el. Grønland: | Udenlandske kvinder | I alt |
|--|--|------------------------|--------|
| "Viede mænd" (dvs. bopæl Dan- mark el. Grønland) | 30.373 | 398 | 30.771 |
| Udenlandske mænd | 421 | . | 421 |
| I alt | 30.794 | 398 | 31.192 |

Omkring 1 1/2 pct. af de viede mænd og kvinder blev således gift med personer med bopæl i udlandet.

En anden fremtrædende årsag til manglende overensstemmelse er, at oplysninger om den ene ægtefælle bliver forsinket.

I 1975 foretoges en stikprøvebaseret opgørelse af årsagerne til manglende match ved samkøringen. Resultatet var følgende:

| | <u>Mænd</u> | <u>Kvinder</u> |
|---|-------------|----------------|
| 1. Ægtefællen bosat i udlandet | 65% | 45% |
| 2. Ægtefællens indberetning forsinket | 25% | 45% |
| 3. Henvisningsnummer mangler fejlagtigt | 5% | 5% |
| 4. Forkert indberetning | 5% | 5% |
| <hr/> | | |
| I alt | 100% | 100% |

Konklusion:

Som følge af folkeregistersystemets indberetningsprocedure kunne oplysninger om vielser kun frembringes gennem samkørsel af to delmaterialer fra CPR.

Samkøringen fungerer, som man måtte vente, næsten perfekt, idet flertallet af de viede, som ikke kan matches, heller ikke skulle kunne matches. På den anden side giver undersøgelsen af de forholdsvis få personer, der ikke matcher, anledning til ressourcekrævende manuelle procedurer.

Eksempel 7: Uddannelsesstatistikken af 1974

1. Statistikkens formål

Statistikken har til formål at beskrive relationen mellem objekttyperne personer, uddannelser og institutioner i Danmark på et givet tidspunkt og i en given periode.

2. Objekttyperne

I statistikken opereres altså med tre objekttyper:

- Personerne afgrænset som de individer, der har søgt eller kan søge en uddannelse i Danmark. Disse individer vil stort set alle være (blive) opført i CPR, og dér beskrives ved personnr. (identificerende variabel), aktuel bopæl, nationalitet mv.
- Uddannelserne afgrænset som den mængde uddannelseselementer (kurser o.l.), der, når alle eller en del heraf er fuldført, medfører en given kompetence (afsluttende uddannelse). Uddannelserne er stort set alle opført i Danmarks Statistiks uddannelsesregister og dér beskrevet ved uddannelseskode (identificerende variabel), indre struktur (ordning, trin og retningsopdeling), varighed, afsluttende uddannelse(r), benævnelser mv.
- Institutionerne afgrænset som den administration og fysiske ramme, inden for hvilken uddannelserne foregår. Institutionerne er stort set alle opført i Danmarks Statistiks (uddannelses)institutionsregister og er dér beskrevet ved institutionsnummer (identificerende variabel), beliggenhed, ejerforhold, ministerielt tilhørsforhold, benævnelse mv.

3. Samkøringsgrundlaget

Uddannelsesinstitutionerne indberetter i tidsafgrænset form alle kombinationer af personnr., udd.nr. og inst.nr., der nyetableres og/eller ophører i en given periode, og for alle kombinationer, der består på et givet tidspunkt (1. oktober), indberettes tillige værdien af statusvariablene ordning, trin og retning.

Indberetningerne opsamles i 5 forskellige registre:

1. omfattende 8. kl. - 3. G samt visse kursusuddannelser,
2. omfattende kurser ved de frie skoler,
3. omfattende lærlingeuddannelserne,
4. omfattende efg-uddannelserne og visse eksamensuddannelser ved

tekniske skoler og handelsskoler samt de sociale, sundheds- og søfartsmæssige uddannelser og endelig

5. omfattende de lange videregående uddannelser.

4. Samkøringerne

4.1. Samkøring (1)

Hvert enkelt af de 5 delregistre samkøres løbende med CPR-registret, uddannelsesregistret og institutionsregistret med det formål at overføre variabelværdier fra disse til en yderligere beskrivelse af de enkelte objekter (bopæl, nationalitet, uddannelsstype, ejerforhold osv.). Lejlighedsvis overføres fra andre registre - fx beskæftigelsesundersøgelserne eller kildekatte-registrene - værdier af andre variable som fx stilling, erhverv, bruttoindkomst osv.

Vurdering: Da man lægger et stort arbejde i at sikre sig elevernes valide personnumre og faktisk opnår det i ca. 99,9 pct. af tilfældene, opnås meget gode resultater af samkøringen med CPR-registret. I forhold til uddannelses- og institutionsregistret er overensstemmelsen 100 pct.

4.2. Samkøring (2)

De færdigdannede delregistre er siden 1974 samlet (objektgruppeintegration) i et integreret elevregister, hvor man har udlagt alle bestående elev-uddannelses-institutionsrelationer pr. 1.10.1974 (og alle, der er afsluttet i løbet af det foregående år). Dette integrerede elevregister ajourføres, efterhånden som nye delregistre (der hver dækker et år) færdiggøres, idet der matches på personnr., udd.nr. og inst.nr., hvis relationen består (burde bestå) i det integrerede elevregister, og der udlægges således løbende dels nye objektrelationer, dels ajourførte værdier af statusvariablene ordning, trin og retning for bestående relationer. Det er formålet med denne samkøring af skabe grundlag for en statistisk beskrivelse af uddannelsesforløbet over tiden og mellem de uddannelsesområder, som de enkelte delregistre dækker.

Vurdering: Samkøringen giver anledning til mismatch i 1-2 pct. af tilfældene (5.-10.000 mismatch), og disse skyldes i overvejende grad, at objektrelationernes identificerende variable er ændret (rettet) af institutionerne. Man er herefter nødt til at foretage en tilsvarende rettelse i det integrerede elevregister, inden ajourføringen glider på plads. Den hermed forbundne arbejdsbyrde

er dog alt i alt til at overse og vil antagelig blive mindre, efterhånden som den nye uddannelsesstatistik indarbejdes.

Eksempel 8: Statistik over døde fra 1973

1. Dødsfaldsstatistikken

Dødsfaldsstatistikken har til formål at belyse dødeligheden blandt Danmarks befolkning.

Der indtræffer ca. 50.000 dødsfald om året, og de variable, der skal belyses, falder i to grupper:

1. De demografiske variable: Køn, alder, civilstand, bopæl, beskæftigelsesforhold mv.
2. De medicinske variable: Dødsårsager, oplysning om obduktion mv.

2. Objektgruppe

Den objektgruppe, dødsfaldsstatistikken skal rette sig imod, er samtlige dødsfald blandt personer med bopæl i Danmark, uanset om dødsfaldet indtræffer i Danmark eller i udlandet.

3. Samkøringsgrundlag

Dødsfaldsstatistikken fremstilles ved kombination af to datamaterialer:

3.1 De originale dødsattester, som udfærdiges af læger ved alle dødsfald, der indtræffer i Danmark.

3.2 Oplysninger fra CPR om alle personer med bopæl i Danmark, der afgår ved døden.

Begge materialer er identificeret ved afdødes personnummer.

For så vidt angår dødsattesterne benyttes kun de attester, der vedrører personer med bopæl i Danmark.

For en del af statistikken objektgruppe - nemlig dødsfald i udlandet blandt danske residerter - findes ingen dødsattester; det drejer sig dog kun om ca. 250 personer om året.

4. Samkøringens formål

Til og med 1972 fremstilledes dødsfaldsstatistikken udelukkende på grundlag af dødsattesterne.

Fra 1973 hentes de demografiske variable hovedsageligt fra CPR, mens de medicinske oplysninger, som jo ikke registreres i CPR, hentes fra dødsattesterne. De to materialer sammenkøres, person for person, med personnummeret som nøgle. Samkørslen foretages månedsvis.

Omlægningen har haft forskellige formål:

4.1. En logisk konsistent befolkningsstatistik

Statistikken over befolkningens størrelse og bevægelser i øvrigt er baseret på oplysninger fra CPR. Ved at lade også dødsfaldsstatistikken hvile på CPR-oplysninger opnår man, at det bliver nøjagtig den samme persongruppe, der belyses i alle dele af befolkningsstatistikken, nemlig personer med bopæl i Danmark. Yderligere opnår man fuldstændig sammenlignelighed mellem fordelinger af henholdsvis dødsfald og den samlede befolkning efter demografiske variable, når disse variable altid hentes fra samme kilde.

4.2. Kvalitetskontrol

Uoverensstemmende oplysninger fra de to kilder fører til en nærmere undersøgelse. Denne kontrol får ikke stor betydning for selve statistikken, men det sammenkoblede materiale udgør tillige det dødskartotek, som føres af Sundhedsstyrelsen.

4.3. Rationalisering

Kodnings- og registreringsarbejdet formindskes ganske væsentligt som følge af samkøringen, fordi det kun er nødvendigt at registrere de medicinske oplysninger fra dødsattesterne.

5. Vurdering af samkøringen

I årene 1973-76 indeholdt CPR-oplysningerne oplysning om 207.237 dødsfald, hvoraf 1.009 indtraf i udlandet. Der var 205.798 dødsattester. Datamængderne og det endelige resultat efter samkøring ses i omstående oversigt:

| | : I hh. til : : CPR : : | : I hh. til : : dødsattester : : | : I hh. til : : samkøring : : |
|-----------------------|-------------------------------|--|-------------------------------------|
| Antal døde i Danmark | 206.228 | 205.798 | 206.050 |
| Antal døde i udlandet | 1.009 | - | 1.009 |
| Antal døde i alt | 207.237 | 205.798 | 207.059 |

De 206.228 CPR-oplysninger om indenlandske dødsfald burde hver svare til én dødsattest. Samkøringen resulterede umiddelbart i, at 99,4% af CPR-oplysningerne kunne kombineres med dødsattester. En manuel sammenligning af de oplysninger, der ikke matchede, viste:

- at yderligere 0,1% af CPR-meddelelserne svarede til dødsattester,
- at 0,2% af CPR-meddelelserne var korrekte, men at dødsattester manglede, og
- at 0,4% af CPR-meddelelserne var forkert placerede med hensyn til dødstidspunkt.

Af de 205.798 dødsattester blev 99,6% kombineret med CPR-oplysninger i samkøringen. I den manuelle gennemgang viste det sig:

- at 0,1% svarede til foreliggende CPR-meddelelser,
- at 0,3% af dødsattesterne var korrekte, men at CPR-meddelelse manglede (hovedsageligt fordi disse meddelelser var indberettet i en anden periode og med forkert dødsdato), og
- at under 1/2 ‰ måtte udgå på grund af forkert dødsdato mv. på attesten.

Konklusion:

Samkøringen fungerer særdeles præcist og problemfrit, idet kun 1/2 pct. af de indenlandske dødsfald kræver en særlig manuel behandling.

En årsag til det gode resultat er, at de anvendte datamaterialer fødes i administrative processer, som for det første er meget sikre på grund af dødsfaldets store juridiske betydning, og som for det andet er meget nært forbundne; således kommer indberetningen til CPR om dødsfaldet fra det kommunale folkeregister, og denne indberetning initieres ved, at dødsattesten forevises i skifteretten.

Eksempel 9: Statistik over levendefødte fra 1973

1. Fødselsstatistikken

Statistikken over levendefødte har til formål at belyse fertiliteten, dvs. den faktiske formering blandt kvinder bosat i Danmark.

Der fødes for tiden ca. 60-70.000 børn om året. De variable, der skal belyses, falder i to grupper:

1. De demografiske variable: Køn, forældres alder og civilstand, bopæl mv.
2. Medicinske variable: Fødselsvægt og -længde, kunstig hjælp mv.

2. Objektgruppe

Den objektgruppe, fødselsstatistikken skal rette sig imod, er samtlige levendefødte børn født af mødre med bopæl i Danmark, uanset om fødslen sker i Danmark eller i udlandet.

3. Samkøringsgrundlag

Fødselsstatistikken fremstilles ved kombination af to materialer:

- 3.1. Jordemoderanmeldelserne¹, som udfærdiges af jordemoder for ethvert levendefødt barn, født i Danmark.
- 3.2. Oplysninger fra CPR om alle levendefødte børn, der fødes i Danmark eller Grønland af kvinder med bopæl i Danmark.

Jordemoderanmeldelserne er ikke systematisk forsynet med barnets personnummer, men begge materialer indeholder, som en sekundær identifikationsvariabel, moderens personnummer.

Ingen af materialerne indeholder oplysninger om hele objektgruppen, eftersom danske kvinders fødsler i udlandet slet ikke registreres.

4. Samkøringens formål

Til og med 1972 fremstilledes fødselsstatistikken udelukkende på grundlag af jordemoderanmeldelserne.

Fra 1973 hentes de demografiske variable hovedsageligt fra CPR, mens de medicinske oplysninger, som jo ikke registreres i CPR, hentes fra jordemoderanmeldelserne. De to materialer sammenkøres, person for person, idet man som nøgle for samkøringen benytter moderens personnum-

¹Fra 1978: Fødselsanmeldelser.

mer kombineret med barnets køn, - hvilke to variable i praksis giver en næsten fuldstændig identifikation. Samkøringen foretages månedsvis.

Omlægningen har haft forskellige formål:

4.1. En logisk konsistent befolkningsstatistik

Statistikken over befolkningens størrelse og bevægelser i øvrigt er baseret på oplysninger fra CPR. Ved at lade også fødselsstatistikken hvile på CPR-oplysninger opnår man, at det bliver den samme persongruppe, der belyses i alle dele af befolkningsstatistikken, nemlig personer med bopæl i Danmark. Yderligere opnår man sammenlignelighed mellem fordelinger af henholdsvis fødsler og den samlede befolkning efter demografiske variable, når disse variable altid hentes fra samme kilde.

4.2. Rationalisering

Kodnings- og registreringsarbejdet formindskes ganske væsentligt som følge af samkøringen, fordi det kun er nødvendigt at registrere de medicinske oplysninger fra jordemoderanmeldelserne.

5. Vurdering af samkøringen

I årene 1973-76 indeholdt CPR-oplysningerne oplysning om 280.598 levendefødte børn, hvoraf 10 fødtes i Grønland. Der registreredes 280.577 jordemoderanmeldelser. Datamængderne og det endelige resultat efter samkøring ses i nedenstående oversigt:

| | : I hh. til : CPR | : I hh. til : jordemoder- : anmeldelser | : I hh. til : samkøring |
|------------------------|----------------------|---|----------------------------|
| Antal fødte i Danmark | 280.588 | 280.577 | 280.550 |
| Antal fødte i Grønland | 10 | . | 10 |
| Antal fødte i udlandet | . | . | . |
| Antal fødte i alt | 280.598 | 280.577 | 280.560 |

De 280.588 CPR-oplysninger om levendefødte i Danmark burde hver svare til én jordemoderanmeldelse. Samkøringen resulterede umiddelbart i, at 99,0% af oplysningerne kunne kombineres med jordemoderanmeldelser, og en manuel gennemgang af de oplysninger, der ikke matchede, viste, at yderligere 1,0% af CPR-meddelelserne svarede til jordemoderanmeldelser, - grunden til manglende matching var hovedsagelig fejlregistrering af anmeldelsens personnummer for moderen. Et helt ubetydeligt antal CPR-oplysninger, 68, måtte udgå, da det drejede sig om fejler dobbeltindberetninger.

Af jordemoderattesterne måtte ligeledes et ubetydeligt antal, nemlig 122, udgå, da mødrene faktisk ikke havde bopæl i Danmark.

Konklusion:

Samkøringen fungerer særdeles præcist og problemfrit, idet kun 1% af de fødte kræver en særlig manuel databehandling, og selv denne sidste procent af de to samkøringsgrundlag er så at sige totalt sammenfaldende.

En årsag til det gode resultat er, at de anvendte datamaterialer fødes i administrative processer, som for det første er meget sikre, og som for det andet er meget nært forbundne. Præsten i fødesognet modtager jordemoderanmeldelsen fra jordemoder, og det er tillige præsten, der indberetter fødslen til folkeregistret, hvorfra meddelelse går videre til CPR.

Eksempel 10: Oplysninger om døde 1970-72

1. Baggrunden

Til en sammenlignende analyse af dødeligheden i forskellige fag- og erhvervsgrupper (jf. eksempel 12) skal man bruge et register over samtlige dødsfald i Danmark 1970-75 med angivelse af dødsårsager; da oplysningerne skal sammenholdes på individniveau med oplysningerne fra folke- og boligtællingen 9. november 1970, er det nødvendigt at hvert dødsfald kan identificeres ved afdødes personnummer.

Siden 1973 har personnummeret været primær identifikation i dødsfaldsstatistikens grundmaterialer (jf. eksempel 8), men før 1973 byggede dødsfaldsstatistikken udelukkende på de originale dødsattester, som dengang ikke var forsynet med personnummer.

2. Objektgruppe

Der skal frembringes et materiale, hvis objekter er samtlige dødsfald 1970-72 blandt personer med bopæl i Danmark, uanset om dødsfaldet indtræffer i Danmark eller i udlandet.

3. Samkøringsgrundlag

Det identificerede materiale frembringes ved kombination af to datamaterialer:

- 3.1. Grundmaterialet for dødsfaldsstatistikken 1970-72, som udgøres af de originale dødsattester; der findes en dødsattest for hvert dødsfald, der har fundet sted i Danmark.
- 3.2. En kopi af CPR's totale indhold den 1. januar 1977; heri findes oplysninger om alle personer, som for tiden er eller som tidligere har stået registreret med bopæl i Danmark. Af særlig interesse er det, at alle afdøde står opført med en særlig "status"-kode og med angivelse af en registreret dødsdato.

Kun CPR indeholder oplysninger om den totale objektgruppe, idet der ikke findes dødsattest for dødsfald i udlandet blandt danske residenter; det drejer sig dog kun om ca. 900 personer i alt i de tre år.

4. Samkøringen og dens formål

De to materialer skal kombineres med det formål at frembringe identificerede dødsfaldsoplysninger.

Der findes ingen fælles primær-identifikation, idet dødsattesterne er identificeret ved et løbenr., mens CPR indeholder personnummret.

De to materialer indeholder en del overlappende variable, som tilsammen identificerer det enkelte dødsfald temmelig præcist:

- 4.1. Fødselsdato
- 4.2. Dødsdato
- 4.3. Køn
- 4.4. Bopælskommune på død tidspunktet

Der er en vis lille risiko for, at flere dødsfald vil vise sammenfald på alle disse variable. Men den alvorligste risiko er dog, at samme dødsfald er registreret forskelligt i de to materialer, dvs. at en eller flere af ovennævnte variable er fejlbehæftede. Risikoen herfor må vokse, når flere variable inddrages i identifikationen, specielt når der ikke findes kontrolsystemer som fx modulus-11 for de enkelte identifikationsvariable.

For at begrænse konsekvenserne af fejl i enkelte variable benyttes en samkøringsprocedure, hvor man først forsøger at matche med alle variable 4.1 - 4.4 som kriterium. Derefter forsøger man successive samkøringer med kriterier, der udgør forskellige delmængder af 4.1 - 4.4, jf. nedenstående tabel:

| Sammen- koblings- nr.: | Sammenkoblingskriterier: | | | | | | | |
|------------------------------|--------------------------|--------------|--------------|--------------|--------------|--------------|-----|------------------|
| | Dødsdato | | | Fødselsdato | | | Køn | Kommune- kode |
| | Dag/Måned/År | Dag/Måned/År | Dag/Måned/År | Dag/Måned/År | Dag/Måned/År | Dag/Måned/År | * | |
| 1 | X | X | X | X | X | X | X | X |
| 2 | X | X | X | X | X | X | X | |
| 3 | | | X | X | X | X | X | X |
| 4 | X | X | X | | | | X | X |

*) incl. årh.

Hver samkøring forsøges kun på den gruppe af objekter, der ikke er blevet forenet i en tidligere samkøring. Før hver samkøring undersøges det, om der i dødsattesterne henholdsvis i CPR findes flere dødsfald, der har samme værdi af samkøringskriterierne, idet sådanne dubletter ikke forsøges samkørt.

De automatiske samkøringsprocesser efterfølges af en manuel undersøgelse af de oplysninger, som ikke kunne matches, idet man som supplerende materiale inddrager gamle lister med navneoplysninger om de døde.

5. Vurdering af samkøring

Det har vist sig, at de fire identifikationsvariable 4.1-4.4 tilsammen udgør en så at sige entydig identifikation, idet kun 0,1% af dødsfaldene måtte betragtes som uidentificerbare i den forstand, at to eller flere dødsfald viste samme værdier af alle fire variable; grunden til sammenfaldet var i næsten alle tilfælde, at det drejede sig om tvillinger, der var døde kort efter fødslen, og det var derfor let ved den efterfølgende manuelle proces at identificere næsten alle disse personer.

Det fremgår i øvrigt af nedenstående oversigt, at der er lidt flere dødsfald i CPR-materialet end i dødsattesterne (0,7%), hvilket svarer godt til omfanget af dødsfald i udlandet.

| | : I CPR | : I dødsat- : testerne : |
|--|---------|-----------------------------|
| Antal dødsfald i alt | 148.435 | 147.437 |
| heraf "uidentificerbare" | 121 | 135 |
| Antal dødsfald til automatisk sam- køring i alt | 148.314 | 147.302 |

Blandt samtlige 148.435 afdøde, der var registreret i CPR, kunne i alt 98,8% matches automatisk, jf. omstående oversigt, der tillige viser, at størsteparten (93,3%) blev matchet allerede i første samkøringsproces.

Efter den manuelle proces resterede 0,9% af samtlige dødsfald, men kun 0,3% kan anslås at have fundet sted i Danmark, således at man burde kunne finde deres dødsattester.

Konklusion:

Selv om primær identifikation mangler, kan samkøring gennemføres med stort held. Det skyldes (som i eksempel 8) den høje datakvalitet og det forhold, at oplysningerne på dødsattesten og i CPR fødes tæt på hinanden i de administrative processer.

| | : : Antal : dødsfald : | : Pct. af : samtlige : dødsfald : (CPR) | : Kumuleret : pct. : matchet ¹ |
|----------------------------|---------------------------------|--|---|
| Dødsfald i alt (CPR) | 148.435 | 100,0 | |
| Heraf: | | | |
| Matchet i samkøring 1 | 138.453 | 93,3 | 93,3 |
| " " " 2 | 2.818 | 1,9 | 95,2 |
| " " " 3 | 3.396 | 2,3 | 97,5 |
| " " " 4 | 1.997 | 1,3 | 98,8 |
| Matchet manuelt | 477 | 0,3 | 99,1 |
| Dødsfald i udland, anslået | 900 | 0,6 | 99,7 ² |
| Rest (ikke matchet) | 394 | 0,3 | . |

Noter:

¹Dvs. andel af samtlige dødsfald, der er matchet i den anførte og de forudgående samkøringsprocesser.

²Andel af samtlige dødsfald, som er matchet eller som ikke skulle kunne matches.

Eksempel 11: Arbejdsstyrketællingen 1975

1. Formål

Formålet med tællingen er at opnå en række variabelværdier - opstillet og defineret af EF - for et udsnit af befolkningen, der er repræsentativ for hele befolkningen. Variabelværdierne er dels demografiske oplysninger som køn, fødselsår, ægteskabelig stilling, statsborgerskab, bopæl, dels oplysninger om de pågældendes normale hovedbeskæftigelse i og uden for erhvervene og deltagelse i erhvervslivet på tællingstidspunktet og året før, og endelig oplysninger om husstandens størrelse og sammensætning.

2. Objektgruppe

Objekterne er personer, der er født den 16. i en lige måned med bopæl i Danmark på tællingstidspunktet, den 15. april 1975.

3. Tællingsteknik

Objekter i tællingen er alle personer, der er født den 16. i en lige måned. For 0-14 årige og for personer på 75 og derover anvendes kun oplysninger, der fremkommer gennem CPR-systemet, idet det forudsættes, at ingen af disse personer er erhvervsmæssigt beskæftiget og tillige at alle personer på 75 år og derover er pensionister. For de 15-74 årige indhentes udover CPR-oplysninger yderligere oplysninger på et spørgeskema udsendt pr. post og for en del af de, der ikke svarer på posthenvendelsen, udsendt med en interviewer. Kun de 15-74 årige, der har besvaret skemaet indgår i de endelige oplysninger til EF.

4. Samkøringsgrundlaget

Indledningsvis foretages en samkøring af

- (A) et udtræk fra CPR, og
- (B) de indkodede skemaværdier med det formål at afgrænse tællingens objektgruppe (AB).

Derefter suppleres dataindholdet for objektgruppen med

1. CPR-status året før, evt. oplysning om indvandring
2. indkodede skemaoplysninger fra tælling året før
3. oplysning om evt. fælleshusholdning
4. familieoplysninger.

Disse fire variabelkombinationer omtales nærmere nedenfor.

5. Samkøringerne

5.1. Objektgruppeafgrænsning

- (A) CPR-udtræk af 16. lige-fødte, der er aktive i CPR i uge 21, 1975. Anvendes som tællingsgrundlag og indeholder visse variabelværdier, nemlig oplysning om køn, fødselsår, ægteskabelig stilling, statsborgerskab, bopæl.

CPR-bestanden i uge 21 antages at svare bedst til den faktiske befolkningsemængde på tællingsdagen, der ligger i uge 16, idet hovedparten af ændringsmeddelelserne indløber til CPR nogle uger efter hændelsestidspunktet.

- (B) Indkodede skemaoplysninger for 15-74 årige. Indeholder variabelværdier vedrørende normal hovedbeskæftigelse og erhvervs-mæssig beskæftigelse mv. på tællingstidspunktet for de 15-74 årige, der har besvaret spørgeskemaet enten pr. post eller ved interview.

Da tællingsskemaet må være de adspurgte i hænde på tællingsdagen, må nødvendigvis bruges et CPR-udtræk, der ligger forud for tællingsdagen.

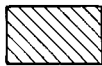
Henvendelse om at besvare skemaet er rettet til personer i alderen 15-74 år, der er aktive i CPR i uge 13, 1975, og som er født på bestemte dage¹.

Sammenfaldet i personbestanden i A og B kan illustreres på følgende måde:

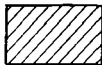
¹ Fødseldagsudsnit er anvendt til at frembringe et repræsentativt udsnit.

Overalt i det følgende anvendes datamaterialer, som omfatter samme fødseldagsudsnit.

| | Fragået CPR | 15-74 årige | CPR uge 21 | Andre |
|----------------------------|----------------|---------------------------|----------------|-------|
| Svarer ikke på skema | | | | |
| Svarer på skema | | | | |
| | | CPR uge 13 15-74 årige | Tilgået CPR | |



Personbestand A



Personbestand B



Personbestand AB

Den endelige afgrænsning af objektgruppen er herefter personbestand AB samt den del af A som ikke er i alderen 15-74 år i uge 21. Der sker en opregning til totalbefolkning med en opregningsfaktor, der er bestemt af forholdet mellem antal objekter og det beregnede befolkningstal på tællingsdagen fordelt efter køn, alder og bopæl.

Da persongruppen "tilgået CPR" er meget lille, må det antages, at dette bortfald er af langt mindre betydning for tællingsresultaterne end det bortfald, der skyldes, at de adspurgte ikke svarer på henvendelsen.

5.2. Variabelkombination

I det følgende omtales 4 sæt variabelkombinationer, som er gennemført med det formål at overføre yderligere oplysninger til personerne i AB, altså den delmængde af den samlede objektgruppe,

som findes i CPR i uge 21, og som har besvaret spørgeskemaet (15-74 årige).

Variabelkombination 1

Data for objektgruppen suppleres ved hjælp af:

- (C) CPR-udtræk af personer i alderen 15-74 på tællingsdatoen i 1975, der er aktive i CPR i uge 14, 1974, dvs. et år tidligere.
 - (D) Udtræk af CPR-indvandrerbånd. Indeholder personer der er registreret som indvandret i perioden april 1974-1975 med oplysning om fraflytningsland.
- (I begge tilfælde medtages kun personer med bestemte fødselsdage, jf. afsnit 5.1. pkt. B).

C og D anvendes til at indsætte variabelværdier for pågældendes bopæl 1 år før tællingen fordelt på lande.

Ved samkøring af (AB) og C vil der være manglende sammenfald i personbestandene af følgende årsager:

- g. 0-15 årige (i 1975) og personer på 76 år og derover (i 1975) findes ikke i (C). Medmindre oplysning om personerne findes på (D), antages de at have haft bopæl i Danmark i 1974.
- h. 16-75 årige, der mellem de tællingstidspunkter er meldt døde eller flyttet til udlandet. Personerne findes i (C) men ikke i (A), og indgår ikke som objekter.
- i. 16-75 årige, der er meldt flyttet til Danmark mellem tællingstidspunkterne. Ved samkørsel af 16-75 årige i (D) (indvandrerbåndet) med persongruppen, der er nævnt under punkt i, skyldes det manglende sammenfald, at tidsrummet mellem de to registerudtræk (uge 14, 1974 og uge 21, 1975) ikke er nøjagtig sammenfaldende med det tidsrum, indvandrerbåndet vedrører. Endvidere bemærkes, at den registrerede indvandringsdato er afgørende for, om hændelsen indgår i D, medens de 2 bestandsopgørelser (registerudtrækkene uge 14, 1974 og uge 21, 1975) ikke er korrigeret for hændelser, der indberettes med forsinkelse. De har derfor en noget diffus tidspunktafgrænsning.

Variabelkombination 2

Data for objektgruppen suppleres ved hjælp af:

- (E) Indkodede skemaoplysninger for 15-74 årige ved beskæftigelsesundersøgelsen 24. april 1974. Omfatter personer i alderen

15-74 år, der er aktive i CPR i uge 14, 1974 og som har besvaret det udsendte tællingsskema i 1974¹.

Samkøring mellem (AB) og (E) har til formål at få variabelværdier for beskæftigelse 1 år før tællingen. Der vil være følgende forskelle i personbestanden:

- k. 0-15 årige (i 1975) og personer på 76 år og derover (i 1975) findes ikke i (E). For disse personer indsættes variabelværdierne svarende til oplysninger "barn" eller "pensionist".
- l. Personer, der i øvrigt findes i (AB) men ikke i (E). Det drejer sig væsentligst om 2 grupper, nemlig personer, der er flyttet til Danmark mellem de to tællingstidspunkter og personer, der ikke har svaret på tællingsskemaet i 1974. For begge grupper bliver den søgte variabelværdi uoplyst.
- m. Personer, der findes i (E) men ikke i (AB). Det drejer sig om personer, der er døde eller flyttet til udlandet mellem de to tællingstidspunkter samt personer, som har svaret på tællingsskemaet i 1974 men ikke i 1975. Personerne indgår ikke som objekter.

På grund af et meget stort antal objekter, hvor den søgte variabelværdi bliver uoplyst, opnås ikke et brugbart resultat, fx ved direkte beregning af erhvervsfordelingen året forud for den aktuelle eller ved analyser af forekomsten af skift i erhvervsmæssig situation. Med hensyn til det sidstnævnte fremkommer et mere brugbart resultat ved at begrænse objektgruppen til de objekter, der har svaret begge år.

Variabelkombination 3

Data for objektgruppen suppleres ved hjælp af:

- (F) Kartotek over fælleshusholdninger. Kartoteket er oprettet i Danmarks Statistik på grundlag af en henvendelse til samtlige kommuner om at meddele navn, adresse og adressekode (vejkode og husnummer) i CPRsystemet for fælleshusholdninger beliggende i kommunen.

Kun fælleshusholdninger, der ikke har adressekode fælles med private husstande helt uden for fælleshusholdningen, er i princippet medtaget i kartoteket. Anvendes til belysning af, hvorvidt de enkelte personer bor i en fælleshusholdning og

¹ Her er benyttet det samme udsnit af fødselsdage som ved arbejdsstyrkeundersøgelsen 1975.

i givet fald, hvilken art fælleshusholdning.

Samkøring af (AB) og (F). Personer i (AB), der har samme adressekode som en af de i kartoteket over fælleshusholdningerne opførte, antages at bo i fælleshusholdningen, og øvrige personer antages at bo i en privat husstand. Da kartoteket ikke rummer samtlige fælleshusholdninger og da tællingen kun omfatter et udsnit af befolkningen, kan ikke anslås et omtrentligt antal personer, der skulle nås ved denne sammenkobling, og det kan ikke forventes, at der er personer på alle fælleshusholdningsadresserne.

Variabelkombination 4

Data for objektgruppen suppleres ved hjælp af:

(G) Udtræk af CPR-familiekerneband, dannet på grundlag af CPR-mastere fra uge 21, 1975. Omfatter familieoverhoveder¹ og disse familiemedlemmer. Indeholder oplysning om de pågældendes køn, fødselsår og stilling i familien (familieoverhoved, ægtefælle, barn).

Samkøring af (AB) og (G) er problemfri, idet såvel personbestanden i (AB) som familieoverhovederne i (G) er uddraget fra CPR-registret på samme tidspunkt og objekter i (AB), der ikke er sammenfaldende med familieoverhoveder i (G) automatisk opfattes som ikke-familieoverhoveder.

¹ Samme udsnit af fødselsdage som arbejdsstyrkeundersøgelsen 1975.

Eksempel 12: Analyse af dødeligheden efter erhverv

1. Analysens formål

Analysen har til formål at sammenligne dødeligheden i forskellige fag-/erhvervsgrupper, idet der skal tages hensyn til dødsrisikoen variation efter køn og alder.

Sammenligningen skal indeholde opdelinger efter dødsårsager.

Det er meningen, at analysens grundmateriale skal kunne benyttes af forskere mv., som ønsker at foretage studier af specielle grupper.

2. Objektgruppe

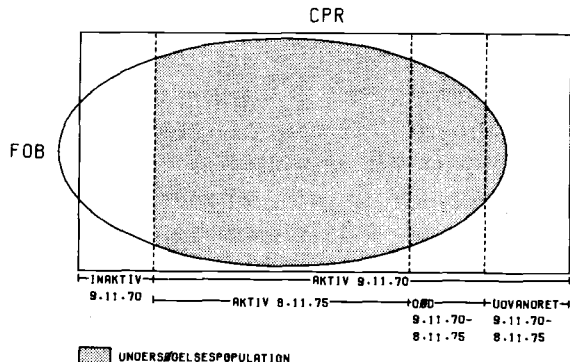
Analysens objektgruppe er personer, der på folketællingsdagen 9.11.70 havde bopæl i Danmark og var 15-74 år gamle (cf. dog afsnit 3). I folketællingsmaterialet haves oplysninger om disse personers beskæftigelsesforhold.

Objektgruppens dødelighed skal analyseres for perioden 9.11.70 til 8.11.75.

3. Analysens princip

I sidste ende skal analysen resultere i beregning af sammenlignelige dødskvotienter for persongrupper defineret ved beskæftigelsesforhold, køn og alder. Aldersspecifikationen og fordelingen efter arbejdsforhold sker i overensstemmelse med situationen på folketællingstidspunktet. En dødskvotient kan således angive, hvor mange ‰ af de ufaglærte mandlige arbejdere ved bygge- og anlægsvirksomhed i alderen 25-29 år ved folketællingen 1970, der dør pr. år i observationsperioden. Kvotienterne skal kunne deles op efter dødsårsager. Selve analysen består derefter i en sammenstilling af kvotienterne.

For at kvotienterne skal give et meningsfyldt udtryk for dødeligheden i de enkelte grupper, må der tages højde for udvandring mv. i perioden 1970-75, idet udvandrede eller forsvundne personer ikke vil være udsat for dødsfald, der registreres i Danmark. Det må også sikres, at der ikke i den ovenfor definerede objektgruppe forekommer personer, som ikke stod tilmeldt CPR pr. folketællingsdagen, idet sådanne personer normalt ikke vil kunne figurere i de eksisterende registre over dødsfald, uanset om de er døde. Konsekvensen heraf er, at objektgruppen så at sige af tekniske eller datakvalitetsmæssige årsager må indsnævres i forhold til afgrænsningen i afsnit 2. Afgrænsningen fremgår af omstående figur.



I øvrigt må bearbejdningen af hensyn til senere specialanvendelser af materialet resultere i en "database", hvor hvert objekt beskrives med hensyn til sin situation på folketællingstidspunktet og med hensyn til sin senere historie, dvs. sin eventuelle udgang af den danske befolkning.

4. Samkøringsgrundlaget

Opgørelserne frembringes ved kombination af tre materialer:

- 4.1. Grundmateriale fra folke- og boligtællingen 1970 (FOB-70) med oplysninger for hver enkelt person.
- 4.2. En kopi af CPR's totale indhold den 1. januar 1977; heri findes oplysninger om alle personer, som for tiden er eller som tidligere har stået registreret med bopæl i Danmark. Personer, som ikke længere har bopæl i Danmark, står opført med en særlig "status"-kode, hvoraf det fremgår, om de er døde, udvandrede eller forsvundne. Desuden findes en dato for deres udgang af befolkningen.
- 4.3. Grundmateriale vedrørende dødsfald 1970-75. Materialet er frembragt ved kombination af dødsattestoplysninger og CPR-oplysninger, jf. eksempel 8 og 10.

Alle tre materialer er identificeret ved personnummer.

5. Samkøringerne, deres formål og vurdering af resultatet

Der foretages to forskellige samkøring, som beskrives i afsnit 5.1 og 5.2:

5.1. FOB-70 kombineres med CPR

Denne samkøring har til formål at afgrænse analysens objektgruppe, jf. afsnit 3 ovenfor, idet de personer ved folketællingen, som ikke burde være der i henhold til CPR's oplysninger, fjernes. Samtidig foretages der en markering af personer, som i henhold til CPR's oplysninger er døde, udvandrede eller forsvundne i løbet af analyseperioden 9.11.70-8.11.75. I samkøringen indgår kun de personer, som var 15-74 år på folketællingstidspunktet.

Indledningsvis har man undersøgt, hvilke personnumre der var registreret mere end én gang i folketællingsmaterialet, og man har fjernet dobbeltregistreringerne, som udgjorde 1 ‰ af den samlede befolkning. Den efterfølgende samkøring viste en høj grad af overensstemmelse mellem de to materialer:

Ud af 3.567.384 personer i FOB-materialet var der kun 1.897 eller 0,5 ‰, som ikke burde have været til stede i henhold til CPR.

En nærmere undersøgelse blev foretaget på en stikprøve blandt fejlene. Den viste, at en forholdsvis ringe andel (35%) af de tilsyneladende ukorrekte FOB-personer var påført folketællingsskemaerne af personerne selv eller af deres husstande.

25% af fejlene var personer, for hvilke man ikke havde kunnet finde et personnummer, hvorfor Danmarks Statistik havde påført et konstrueret nummer. De sidste 40% af fejlene var personnumre, der var påført af kommuner; halvdelen af disse personnumre tilhørte udlændinge, der i mange tilfælde kan antages at være flyttet til deres hjemland, uden at kommunen på folketællingstidspunktet har været vidende herom.

5.2. Analysens objektgruppe kombineres med dødsfaldsoplysninger

Formålet med denne samkøring er at komplettere forløbsbeskrivelsen for hver person i analysens objektgruppe. Der er allerede under foregående samkøring etableret en foreløbig forløbsbeskrivelse, idet personer, der er døde, udvandrede eller forsvundne, er blevet markeret. Den her omtalte proces bevirker dels en afstemning af dødsfaldsoplysningerne med den officielle statistik, dels en komplettering med dødsårsagsoplysninger.

Med personnummeret som nøgle samkøres objektgruppen med dødsfaldsoplysningerne 1970-75.

Ved samkøringen, som omfattede i alt 3.565.487 personer og 301.493 dødsfald, blev antallet af konstaterede fejl meget beskedent:

1. Første type fejl er, at personen i henhold til CPR skulle være død 1970-75, men at dette ikke fremgår af dødsfaldsmaterialet; af denne type fandtes 80 personer.
2. Anden type fejl er, at personen i henhold til CPR ikke skulle være død i perioden, men at der alligevel findes en dødsfaldsoplysning; af denne type fandtes 18 personer.

Hertil kommer en række tilfælde, hvor man har dødsfaldsoplysning om en person, der ikke findes registreret i objektgruppen. Der fandtes 1.184 dødsfald af denne art, men der er ikke nødvendigvis tale om fejl, idet det kan være - og sandsynligvis er - personer, der er indvandret efter folketællingstidspunktet, og som derefter er døde. Der er ikke gennemført en egentlig undersøgelse heraf.

Hvis der faktisk måtte være tale om mangler i folketællingsmaterialet, ville de dog rimeligvis ikke bevirke skævheder i analysens konklusioner, men blot føre til en mindre optimistisk vurdering af materialets sikkerhed.

6. Konklusion

Der er her tale om en forløbsanalyse, hvor man følger en bestemt objektgruppes bevægelser gennem en given tidsperiode. Sådanne analyser kan kun gennemføres ved anvendelse af samkøringsteknik, men de stiller samtidig betydelige krav til datas kvalitet og sikkerhed. Uoverensstemmelser i identifikationer og tidsangivelser vil give fejlagtige konklusioner med hensyn til bevægelsernes intensitet.

I det foreliggende tilfælde synes datakvaliteten at være meget høj, hvilket også var at vente i betragtning af folkeregistreringens og specielt dødsfaldsregistreringens betydning for administrationen. Det har formentlig været af afgørende betydning, at de kommunale myndigheder i 1970-71 gennemførte en omfattende afstemning mellem folketællingens skemamateriale og indholdet i de kommunale folkeregistre.

Eksempel 13: Dannelse af oplysning om udlejningsforhold i bygnings- og boligregistret

1. Baggrunden

I forbindelse med boligstatistik har man brug for at kunne klassificere boligerne efter, om de er:

1. Udlejet (beboet, men ikke af ejeren)
2. Benyttet af ejeren
3. Ikke benyttet.

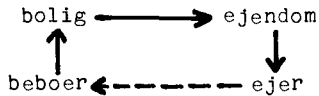
Oplysningen skal, foruden til statistik, bruges administrativt ved beregning af lejeværdi for ejerbenyttede enheder i en ejendom og i beregninger af bloktilskud til kommunerne. Samkøringen sker derfor ikke hos Danmarks Statistik, men i Boligministeriets regi.

2. Objektgruppe

Den objektgruppe, man interesserer sig for, er samtlige boliger i Danmark.

For at frembringe den ønskede oplysning, må man betragte boligernes relationer til nogle andre objekttyper: Ejendomme, ejere og beboere.

Relationerne kan opstilles således:



En bolig er beliggende i en given ejendom, som ejes af én eller flere personer, et selskab eller det offentlige.

Boligen kan være beboet (dvs. at én eller flere personer har bopæl i den pågældende bolig og kun den) og én eller flere af beboerne kan være ejer af den ejendom, hvori boligen er beliggende.

3. Samkøringsgrundlag

- a. Bygnings- og boligregistret (BBR); omfatter alle boliger beliggende i Danmark med oplysning om relation til ejendom.
- b. Det fælleskommunale Ejendomsdatasystem (ESR); omfatter alle ejendomme beliggende i Danmark med oplysning om relation til ejere.
- c. Det Centrale Personregister (CPR)¹.

4. Samkøringen

Udgangspunktet er bestanden af boliger, som er registreret i BBR. For hver af disse har man i BBR registreret et ejendomsnr., som henviser til ejendomsregistreringerne i ESR. De to registre samkøres, hvorved man til hver bolig kan knytte identifikation af ejendommens ejer(e) fra ESR. Identifikationerne af ejerne består i personnr. for private ejere og virksomhedsnumre for andre.

Forbindelsen mellem BBR og CPR skabes af en fælles, entydig adresseidentifikation, som i CPR angiver personernes bopælsadresse. Denne forbindelse udnyttes til at samkøre de to registre, hvorved man til hver bolig kan tilknytte oplysning om personnumre for eventuelle beboere.

Hvis man herefter kan konstatere, at mindst én person optræder som både ejer og beboer, betragtes boligen som "benyttet af ejer". Boliger, der ikke bebos af en ejer, betragtes som "udlejet", hvis CPR viser persontilmelding og ellers som "ikke benyttet".

Ved samkøringen af de tre registre foretages der ingen tidsafgrænsning; registrene samkøres altså med det indhold, de har på samkøringsdatoen. Dette indebærer, at der ikke tages højde for ændringer, der indberettes efter samkøringsdatoen, men som er indtruffet i perioden før denne dato. Samkøringen foretages én gang om året; oprindeligt var det forudsat, at samkøringen skulle ske omkring årsskiftet, men både i 1980 og i 1981 fandt den sted i marts.

5. Vurdering af samkøring

5.1. Manglende tidsafgrænsning

Den manglende tidsafgrænsning af registrene resulterer i en række fejl og ukorrektheder, jf. nedenstående eksempler. Fejlomfanget er herudover betinget af ajourføringsrytmen for registrene. CPR ajourføres dagligt, mens BBR og ESR opdateres månedligt.

A. En ejerlejlighed er solgt i januar, men ejerskiftet er på samkøringstidspunktet ikke registreret i ESR. Den nye ejer er imidlertid flyttet ind i ejerlejligheden, og flytningen er registreret i RFR:

¹I praksis benyttes den version af CPR, som kommunerne fører kaldet det regionale folkeregister.

Virkelighed 15. marts 1980

Person B er ejer
 Person B bebor boligen
 Boligen er beboet af ejer

Registerindhold 15. marts 1980

ESR: Person A er ejer
 CPR: Person B bebor boligen
 Resultat af samkøringen:
 Boligen er udlejet

B. En udlejningsbolig er fraflyttet af de tidligere beboere, og en anden familie er flyttet ind. Fraflytningen er registreret i RFR på samkøringstidspunktet, men tilflytningen er ikke registreret.

Virkelighed 15. marts 1980

Familie A er fraflyttet
 Familie B er tilflyttet
 Boligen er udlejet

Registerindhold 15. marts 1980

ESR: Boligselskab er ejer
 CPR: Ingen tilmeldt
 Resultat af samkøringen:
 Boligen er ikke benyttet

C. En udlejningsbolig er fraflyttet af de tidligere beboere og de nye beboere er ikke flyttet ind. Fraflytningen er ikke registreret i RFR.

Virkelighed 15. marts 1980

Boligen er ikke beboet

Registerindhold 15. marts 1980

ESR: Boligselskab er ejer
 CPR: Familie A tilmeldt på adressen
 Resultat af samkøringen:
 Boligen er udlejet

D. Et parcelhus er færdigbygget i januar 1980, men er ikke overført fra BBR's ændringsregister til stamregistret. Parcelhuset er ikke registreret i ESR, men ejerfamilien er flyttet ind og er registreret på adressen i RFR.

Virkelighed 15. marts 1980

Familie A har færdigbygget nyt parcelhus og er flyttet ind.

Registerindhold 15. marts 1980

ESR: Ingen registrering
 CPR: Familie A tilmeldt på adressen
 Resultat af samkøringen:
 Ingen registrering.

5.2. Identifikationssystemets kvalitet

Som nævnt udnyttes den entydige boligadresse (adressekoden) som sammenkædningsgrundlag mellem BBR og CPR. Man véd imidlertid, at adressen ikke er tilstrækkelig sikker, idet den kan være registreret mere eller mindre afvigende i de to registre, specielt i etageejendomme.

I sådanne tilfælde vil en person blive opfattet som boende i en forkert bolig, og en bolig kan fejlagtigt blive opfattet som ubeboet.

5.3. Talmæssig belysning

Der er ikke foretaget forsøg på egentlig måling af omfanget af de omtalte fejl. En vis idé om, hvad den manglende tidsafgrænsning betyder, kan dog fås ved at sammenholde samkøringsresultatet med Danmarks Statistiks boligtællinger pr. 1. januar, som bygger på CPR og BBR, og som giver en belysning af beboerne (husstandens) boligforhold.

I boligtællingen er der foretaget en tidsafgrænsning af både person- og boligoplysningerne. For CPR's vedkommende er der taget højde for ændringer i perioden før 1. januar, der indberettes til registret senest 40 dage efter den 1. januar. Ved tidsafgrænsningen af BBR anvendes to kopier af registret idet kopien pr. 1. januar suppleres med en kopi fra 1. april.

Boligtællingen resulterer i oplysninger om boliger, hvor der er tilmeldt personer, og om boliger, hvor der ikke er tilmeldt personer. Nu er der som nævnt ikke fuldstændig overensstemmelse mellem adresseidentifikationerne i de to registre, så den sidstnævnte gruppe - boliger uden persontilmeldning i boligtællingen - vil omfatte både reelt tomme boliger og beboede boliger, hvis adresse ikke findes i CPR på grund af manglende adresseensretning; de nævnte tomme boliger behøver desuden ikke at være ubenyttede, idet de kan være "sekundærboliger" (fx for folketingsmedlemmer, der normalt har bopæl i Jylland), eller de kan være beboet af personer, der ikke er tilmeldt dansk folkeregister (fx diplomater). Problemet med den manglende fuldstændige adresseoverensstemmelse gør sig på samme måde gældende ved samkøringen af CPR, BBR og ESR til dannelsen af udlejningsforholdskoden.

Konsekvenserne af den manglende tidsafgrænsning ved dannelsen af kodeværdierne for udlejningsforhold kan illustreres ved at sammenholde samkøringens resultater med resultaterne fra boligtællingen, idet samkøringstidspunkterne ligger rimelig tæt op ad hinanden. I nedenstående tabel er vist forskellene mellem de to samkøringer foråret 1980.

| Udlejningsforhold i BBR 1980 | | | | | |
|--------------------------------|---------------|------------------|---------------|-----------------------|---------|
| Boligtælling | Udlejet | Benyttet af ejer | Ikke benyttet | Uoplyst ¹⁾ | Ialt |
| 1.1.80: | antal boliger | | | | |
| | a | b | c | d | e |
| Boliger med personer tilmeldt | 862862 | 1092679 | 19571 | 6211 | 1981323 |
| | f | g | h | i | j |
| Boliger uden personer tilmeldt | 16839 | 5882 | 102815 | 2534 | 128070 |
| | k | l | m | n | o |
| Egentlige boliger i alt | 879701 | 1098561 | 122386 | 8745 | 2109393 |

¹⁾ Boliger som ikke indgik i BBR på tidspunktet for dannelse af udlejningsforholdsoplysningen.

Det ses af tabellen, at 2.058.356 (celle a+b+h) eller 97,6% af alle egentlige boliger havde samme kobling mellem boliger og personer ved de to opgørelsesmetoder. 19.571 boliger (celle c), som var beboede ifølge boligtællingen, var ubenyttede ifølge kodeværdien i feltet for udlejningsforhold. Forklaringen herpå er bl.a. flytninger, der ikke var registreret i CPR. 8.745 boliger (celle d+i) fandtes ikke i BBR ved dannelsen af udlejningsforholdskoden, hvilket hovedsagelig må skyldes forsinket overførsel af nybyggede boliger fra BBR's ændringsregister til stamregistret. 16.839 boliger (celle f) var udlejede og 5.882 boliger (celle g) var benyttet af ejer ifølge koden for udlejningsforhold, mens de ved boligtællingen fremstod som tomme. Forklaringen herpå kan bl.a. være "flyttetomhed", der ikke er registreret i CPR.

Bilag 3: Statistikregistrene på person-
statistikområdet og deres kilder

Tabellen på næste side giver en oversigt over statistikregistrene på personstatistikområdet og over, hvordan data fra de forskellige administrative registre kanaliseres ind i disse registre. De to sidste kolonner i tabellen viser datastrømme til statistikregistrene fra andre kilder end de administrative registre.

I mange tilfælde er gangen den, at et bestemt statistikregister modtager administrative grundoplysninger og underkaster dem en statistisk grundbearbejdning, hvorefter de bearbejdede oplysninger passerer videre til andre statistikregistre. Herved sikres konsistensen mellem statistikregistrene. Som eksempel herpå kan nævnes arbejdsklassifikationsmodulet, som grundbearbejder en række beskæftigelsesdata og afstemmer dem overfor hinanden. Resultatet heraf står til rådighed for opgørelser på alle personstatistikområder, men det overføres direkte fra arbejdsklassifikationsmodulet til folke- og bolig-tællingsregistret (og herigennem også til erhverv/ dødelighedsstatistikregistret), indkomststatistikregistret og beskæftigelsesstatistikregistret, idet arbejdsklassifikationer på disse statistikområder er nødvendige for den løbende statistik.

Disse forhold er i tabellen antydnet ved markeringer med + og - ; + betyder, at oplysninger fra den pågældende kilde direkte anvendes ved opdateringen af det pågældende statistikregister, mens - betyder, at der ved opdateringen af det pågældende statistikregister er anvendt data fra et andet statistikregister, som bygger på oplysninger fra det pågældende administrative register.

BILAG 3 (fortsat)

Tabel. STATISTIKREGISTRENE PÅ PERSONSTATISTIKOMRÅDET OG DERES KILDER.

| | ADMINISTRATIVT FØDEREGISTER | | | | | | | | | | | ANDRE KILDER | | | | | |
|---|-----------------------------|-----|-------|-------|-------|------------------|------------------|-----------------------|---------------------------------------|------------------------------|-------------------------|---------------------------|---------------------------|--|--|--|-----|
| | CPR | BBR | COR-S | CSR-P | CSR-L | Erhvervsregister | Off. lønregister | Sømandsskatteregister | Det centrale kriminalløstrageregister | Økonomi- og indtægtsregister | Administrativt register | Statistisk grundmateriale | Statistisk dataindsamling | | | | |
| STATISTIKREGISTRE: | | | | | | | | | | | | | | | | | |
| 1. Befolkningsstatistikregistret | + | | | | | | | | | | | | | | | | |
| 2. Medicinsk fødsels- og dødsfaldsstatistikregister | + | | | | | | | | | | | | | | | | |
| 3. Folke- og boligtilingsregistret | + | + | | | | | | | | | | | | | | | |
| 4. Folke- og boligtilingsregistret for Grønland | - | | | | | | | | | | | | | | | | |
| 5. Erhverv/dødelighedsstatistikregistret .. | + | | | | | | | | | | | | | | | | |
| 6. Indkomststatistikregistret | - | | + | | + | | | | | | | | | | | | |
| 7. Økonomisk statistisk register for Grønland .. | - | | | | + | | | | | | | | | | | | |
| B. Beskæftigelsesstatistikregistret | - | | + | | | | | | | | | | | | | | |
| 9. Beskæftigelsesundersøgelsesregistret ... | + | | | | | | | | | | | | | | | | |
| 10. Ungdomsbeskæftigelsesundersøgelsesregistret | + | | | | | | | | | | | | | | | | |
| 11. Arbejdsstyrkeundersøgelsesregistret | + | | | | | | | | | | | | | | | | |
| 12. Løn- og personalestatistikregister | | | | | | | + | | | | | | | | | | |
| 13. Arbejdsklassifikationsmodulet | - | | + | + | + | + | | | | | | | | | | | |
| 14. Arbejdspladsstatistikregistret | | | + | | | + | | | | | | | | | | | (+) |
| 15. Arbejdsløshedsstatistikregister | | | | | | | | | + | | | | | | | | |
| 16. Selvangivelsesstatistikregistret | - | | | | | | | | | | | | | | | | |
| 17. Uddannelsesstatistikregistret | + | | | | | | | | | | | | | | | | |
| 18. Statistikregister vedr. langvarigt uddannedes beskæftigelse | - | | | | | | | | | | | | | | | | |
| 19. Uddannelsesklassifikationsmodulet | - | | | | | | | | | | | | | | | | |
| 20. Bistandslovsstatistikregistret | - | | | | | | | | | | | | | | | | |
| 21. Kriminalstatistikregistret | | | | | | | | | + | | | | | | | | |
| 22. Fødselsuhedsstatistikregistret | | | | | | | | | | | | | | | | | |
| 23. Huslejepanelets register | (+) | (+) | | | | | | | | | | | | | | | |
| 24. Udsnitsarkivet | - | | | | | | | | | | | | | | | | |
| 25. Medicinsk dødsfaldsstatistikregister for Grønland | + | | | | | | | | | | | | | | | | |
| 26. Omnibusundersøgelsesregistret | + | | | | | | | | | | | | | | | | |
| 27. Forbrugsundersøgelsesregistret | + | | | | | | | | | | | | | | | | |

Anm.:

+ betyder, at kilden anvendes direkte.

- betyder, at et andet statistikregister, som bygger på den pågældende kilde, er anvendt.

Bilag 4: De enkelte statistikregistre på
personstatistikområdet

I dette bilag gives en summarisk beskrivelse af registrene på personstatistikområdet. Løbenumrene svarer til de i bilag 3 anvendte.

Hvert register omtales i et særligt afsnit med flg. punkter:

- Formål
- Definition af personkreds
- Dækning (stikprøve/totaldækning)
- Datatyper
- Delregistre (eventuelt)
- Tidsmæssig dækning
- Datagrundlag
- Opdateringsfrekvens

Med hensyn til formålsspecifikationen bemærkes, at det kun er det umiddelbare formål, nemlig fremstillingen af en række specificerede statistikker, der omtales. Statistikkerne tjener igen forskellige formål indenfor administrationen, planlægning, grundforskning og undervisning.

Oplysningerne i et register vil desuden ofte kunne finde anvendelse som baggrundsoplysninger indenfor et andet statistikområde.

1. Befolkningsstatistikregistret

Formål

1. At danne grundlag for løbende befolkningsstatistik, dvs. opgørelser over befolkningens størrelse og sammensætning (herunder familistatistik) og om de bevægelser, befolkningen er underkastet som følge af fødsler, dødsfald, flytninger, vielser, skilsmisser, adoptioner, ind- og udvandring.
2. Det er tillige registrets formål at danne en fælles referenceramme for samtlige andre personstatistikområder, således at fx familiebegreber og ensartede områdeinddelinger kan benyttes på disse statistikområder.
3. Endelig skal registret kunne benyttes til udtrækning af stikprøver (herunder stratificerede stikprøver) til brug ved spørgeskemaundersøgelser; adresseringen sker da i CPR.

Definition af personkreds

Registret retter sig mod den samlede befolkning, der har bopæl i Danmark og Grønland.

Dækning

Total.

Datatyper

Registret indeholder oplysninger om dels befolkningen på givne tidspunkter, dels om ændringerne i befolkningen. Dataindholdet er almindelige folkeregisterdata såsom køn, alder, civilstand, bopæl, nationalitet og familiehenvi- sning. For bevægelserne findes oplysning om hændelsestype og -tidspunkt, og personens situation karakteriseres med de nævnte data før og efter hændelsen. Personnummer er identifikation.

Delregistre

1. Statusregistret, som indeholder oplysning om befolkningen på givne tidspunkter. Statusregistret opbevares i to versioner, en personnummersorteret og en adressesorteret.
2. Bevægelsesregistret, som akkumulerer bevægelsesoplysningerne.
3. Områdekataloget, som indeholder nøgle mellem adresser og områder.
4. Genvandringsregistret, som akkumulerer oplysninger om gendvandring blandt indvandrede og genindvandring blandt udvandrede.

Tidsmæssig dækning

Registret dækker årene siden 1975. Dog findes bevægelsesoplysningerne tilbage til 1971.

Datagrundlag

Tilstandsudtræk og ændringsudtræk fra CPR's personregister. Tilstandsudtræk fra CPR's vejregister.

Opdateringsfrekvens

1. Statusregistret og områdekataloget opdateres (fornys) én gang årligt pr. 1. januar.
2. Bevægelsesregistret suppleres én gang ugentlig med ændringer fra CPR.
3. Statusregistret føres efter behov frem til vilkårlige datoer ved hjælp af bevægelsesregistret.
4. Genvandringsregistret opdateres årligt på grundlag af bevægelsesregistret.

2. Medicinsk fødsels- og dødsfaldsstatistikregister

Formål

At danne grundlag for statistik ang. fødsler og dødsfald.

Definition af personkreds

Registret retter sig mod den samlede befolkning, der har bopæl i Danmark. I registret optages oplysning om fødsler (herunder dødfødsler) samt dødsfald i denne befolkning.

Dækning

Total.

Datatyper

Registret indeholder oplysning om:

1. Personernes personnummer og almindelige folkeregisteroplysninger (bopæl, civilstand, nationalitet) på hændelsestidspunktet.
2. For fødslernes vedkommende tilsvarende oplysninger om forældrene.
3. Oplysninger om hændelsens tidspunkt og om medicinske forhold i forbindelse med hændelsen (fx for dødsfald om årsag og for fødsler om kunstig igangsættelse).

Delregistre

1. Levendefødtregister
2. Dødfødtregister
3. Døderegister
4. Spædbørnsdøderegister (heri indgår oplysninger om omstændighederne ved såvel fødsel som død for børn, der dør før 1-årsfødselsdagen)

Tidsmæssig dækning

| | | | | |
|-----------------------|--------|----------|-------|------|
| Levendefødtregister | dækker | perioden | siden | 1973 |
| Dødfødtregister | " | " | " | 1971 |
| Døderegister | " | " | " | 1970 |
| Spædbørnsdøderegister | " | " | " | 1973 |

Datagrundlag

Oplysningerne i levendefødte- og døderegister tilvejebringes ved samkøring af oplysninger om fødte og døde personer fra CPR med oplysning om hændelsens medicinske aspekter, som stammer fra hhv. fødselsanmeldelser og dødsattester.

Spædbørnsdøderregistret er baseret på de to førstnævnte delregistre.

Dødfødtregisteret er udelukkende baseret på fødsels- og dødsattesterne.

Opdateringsfrekvens

Levendefødte- og døderegister opdateres månedligt. Dødfødte- og spædbørnsdøderegister opdateres årligt.

Særlige bemærkninger

Registret produceres og opdateres i Danmarks Statistik i samme proces som Sundhedsstyrelsens medicinske fødsels- og dødsfaldsregister. Formidling af personoplysninger til anvendelse i medicinsk forskning m.v. udenfor de to institutioner sker udelukkende fra Sundhedsstyrelsens register.

3. Folke- og bolig-tællingsregistret

Formål

At danne grundlag for en bred statistisk belysning af befolkningens beskæftigelsesforhold, uddannelsesforhold, familieforhold, boligforhold og pendling på givne tidspunkter med ca. 5 års mellemrum.

Definition af personkreds

Registret omfatter samtlige personer med bopæl i Danmark.

Dækning

Total.

Datatyper

Identifikation er personnummer. Derudover findes oplysninger om beskæftigelsesforhold, uddannelsesforhold, familieforhold, bopæl og boligforhold.

Tidsmæssig dækning

Registret går tilbage til 1970.

Datagrundlag

Oplysningerne blev pr. 1970 indhentet direkte fra borgerne ved hjælp af spørgeskemaer; ved de senere opgørelsesterminer (1976 og 1981) indsamles oplysningerne fra administrative registre. Opdateringen sker da ved samkøring af oplysningerne i en række statistikregistre:

1. Befolkningsstatistikregistret, (lb.nr. 1)
2. Arbejdsklassifikationsmodulet, (lb.nr. 13)
3. Arbejdspladsstatistikregistret, (lb.nr. 14)
4. Uddannelsesklassifikationsmodulet, (lb.nr. 19)

Desuden anvendes data fra bygnings- og boligregistret (BBR) til belysning af boligforhold.

Opdateringsfrekvens

Registret opdateres ca. hvert femte år.

4. Folke- og boligtællingsregistret for Grønland

Formål

At danne grundlag for en bred statistisk belysning af den i Grønland bosatte befolknings erhvervsforhold, uddannelsesforhold, familieforhold og boligforhold.

Definition af personkreds

Registret omfatter samtlige personer med bopæl i Grønland.

Dækning

Total.

Datatyper

Identifikation er personnummer. Derudover findes oplysning om beskæftigelsesforhold, uddannelsesforhold, familieforhold, bopæl og boligforhold.

Tidsmæssig dækning

Registret går tilbage til 1970.

Datagrundlag

Oplysningerne indhentes direkte fra borgerne ved hjælp af spørgeskemaer, der udsendes til kommunerne, som omdeler skemaerne til de enkelte husstande.

Opdateringsfrekvens

Registret opdateres ca. hvert femte år.

5. Erhverv/dødelighedsstatistikregistret

Formål

At danne grundlag for analyser af sammenhængen mellem beskæftigelsesforhold og dødsrisiko. Analyserne kan være generelle eller selektive mht. personkreds, dødsårsagsgrupper mv.

Definition af personkreds

Registret omhandler personer, der var 15-74 år og havde bopæl i Danmark ved folketællingen 1970.

Dækning

Total.

Datatyper

1. Oplysninger om personens forhold ved folke- og bolig­­tællingen 1970, herunder om beskæftigelsesforhold, uddannelsesforhold, familieforhold og boligforhold. Identifikation: Personnummer.
2. Oplysninger om personens beskæftigelsesforhold ved senere folketællinger.
3. Oplysning om tidspunkt og årsag (død eller udvandring) for personens evt. afgang fra befolkningen i perioden efter folketællingen. I tilfælde af død findes tillige medicinske oplysninger om dødsfaldet, fx om årsag.

Tidsmæssig dækning

Registret går tilbage til 9.11.1970.

Datagrundlag

1. Folke- og bolig­­tællingsregistret (lb.nr. 3)
2. Medicinsk fødsels- og dødsfaldsstatistikregister (lb.nr. 2)
3. CPR

Opdateringsfrekvens

Registret opdateres hvert femte år (sidste gang 1975) med oplysninger om afgang af personer i den forløbne periode samt med folke- og bolig­­tællingsregistrets oplysninger om beskæftigelsesforhold ved den nye femårsperiodes begyndelse.

6. Indkomststatistikregistret

Formål

1. Registret danner grundlag for statistiske opgørelser af personlige indkomster, fradrag og formuer. Det anvendes dels personorienteret og dels familieorienteret. Sigtet er at belyse indkomstforholdene mv. i de enkelte år samt udviklingen fra år til år for befolkningsgrupper afgrænset efter forskellige beskæftigelsesmæssige og geografiske kriterier.
2. Endvidere er registret grundlag for supplerende opgørelser vedrørende personskatstatistik og kirkestatistik.

Definition af personkredsen

Registret omfatter personer med bopæl i Danmark samt personer, der i øvrigt skal svare indkomst- eller formueskat til staten.

Dækning

Total.

Datatyper

I registret indgår selvangivelsens hovedposter samt visse andre grunddata og beregnede data fra ligningsprocessen, specifikation af enkelte beløb svarende til oplysningssedlens indhold, arbejdsklassificerende data og folkeregisterdata.

Delregistre

1. Personorienteret indkomstregister
2. Familieorienteret indkomstregister

Tidsmæssig dækning

Delregister 1 dækker årene fra 1976, og delregister 2 dannes fra 1977.

Datagrundlag

Dele af det centrale skatteyderregisters ligningsdel (CSR-L), det centrale oplysningsseddelregister (COR-S), befolkningsstatistikregistret (lb.nr. 1) samt arbejdsklassifikationsmodulet (lb. nr. 13).

Opdateringsfrekvens

Arlig.

7. Økonomisk statistisk register for Grønland

Formål

At danne grundlag for den økonomiske statistik for Grønland.

Definition af personkreds

Registret omfatter personer, der har bopæl i Grønland, er skattepligtige eller er ejere af et i Grønland indregistreret fartøj.

Dækning

Total.

Datatyper

Registret indeholder oplysninger om demografiske forhold, indkomst- og formueforhold. For selvstændige erhvervsdrivende findes oplysning om virksomhed og evt. fartøj.

Identifikationer er personnummer, arbejdsgivernummer og fartøjsregisternummer.

Tidsmæssig dækning

Fra 1976.

Datagrundlag

De fleste oplysninger er blanketindsamlede oplysninger fra de grønlandske myndigheder. Hertil kommer oplysninger fra Statsskattedirektoratets skatteyderregister, ligningsdelen og fra befolkningsstatistikregistret (lb.nr. 1).

Opdateringsfrekvens

Registret opdateres årligt, dog opdateres oplysningerne om fartøjsejere kvartalsvis.

8. Beskæftigelsesstatistikregistret

Formål

Registret danner grundlag for statistiske opgørelser af beskæftigelsen efter erhverv og stilling på en given dag i året. I forhold til beskæftigelsesundersøgelserne sigter registret i særlig grad mod regionale opgørelser.

Definition af personkreds

Samtlige personer med bopæl i landet på en given dag i året.

Dækning

Total.

Datatyper

I registret indgår oplysning om stilling og erhverv, arbejdsstedets beliggenhed samt visse folkeregisterdata som køn, alder, civilstand, bopæl. Personnummer er identifikation.

Tidsmæssig dækning

Fra 1977.

Datagrundlag

Oplysningsseddelregistret (COR-S), arbejdspladsstatistikregistret (1b.nr. 14), befolkningsstatistikregistret (1b.nr. 1), arbejdsklassifikationsmodulet (1b.nr. 13).

Opdateringsfrekvens

Årlig.

9. Beskæftigelsesundersøgelserregistret

Formål

1. At opgøre arbejdsstyrkens størrelse og sammensætning.
2. At danne grundlag for opgørelse af bevægelserne i arbejdsstyrkens størrelse og sammensætning fra én tælling til den næste.

Definition af personkreds

Beskæftigelsestællingen omfatter personer i alderen 15-74 år på tællingstidspunktet, som har bopæl i Danmark (ekskl. Grønland).

Dækning

Den enkelte tælling omfatter et udsnit på mellem 1,6 og 4,8 pct. af ovennævnte personkreds. Udvælgelsen er bestemt ud fra fødselsdatoer, der udskiftes efter et fastlagt mønster.

Datatyper

Oplysning om beskæftigelsesforhold mv. på tællingsdagen samt civilstand og bopæl. Personnummer er identifikation.

Tidsmæssig dækning

Første tælling afholdtes i 1967. Registrerede oplysninger er bevaret fra 1969, delvis i udsnitsarkivet.

Datagrundlag

1. Personligt besvarede postspørgeskemaer.
2. Det centrale personregister.

Opdateringsfrekvens

Årligt, i 1973 og 1974 dog to gange årligt.

10. Ungdomsbeskæftigelsesundersøgelserregistret

Formål

At belyse de unges beskæftigelses- og ledighedsforhold i relation til deres uddannelsessituation.

Definition af personkreds

Ungdomsbeskæftigelsesundersøgelserne har hidtil omfattet personer i alderen 15-24 og 15-27 år i Danmark ekskl. Grønland.

Dækning

Tællingerne har omfattet hhv. 4,8 og 5,9 pct. af ovennævnte personkredse. Udvælgelsen har været bestemt ud fra fødselsdatoer med varierende regionale udsnitsandele.

Datatyper

Oplysning om beskæftigelsesforhold mv. på tællingsdagen, uddannelse samt oplysninger om lediges arbejdssøgen. Endvidere oplysning om civilstand. Personnummer er identifikation.

Tidsmæssig dækning

Første tælling 1975.

Datagrundlag

1. Personligt besvarede postspørgeskemaer.
2. Det centrale personregister.

Opdateringsfrekvens

Hidtil hvert tredje år.

11. Arbejdsstyrkeundersøgelserregistret

Formål

1. At opgøre arbejdsstyrkens størrelse og sammensætning samt at belyse beskæftigelses- og ledighedsforhold.
2. I tilknytning hertil at belyse særlige forhold, der vedrører deltagelse på arbejdsmarkedet, ud fra retningslinier fastlagt i det harmoniserede program for arbejdskraftopgørelser inden for de europæiske fællesskaber.

Definition af personkreds

Arbejdsstyrketællingen omfatter personer, som på tællingstidspunktet har bopæl i Danmark ekskl. Grønland. Spørgeskemaoplysninger indhentes kun for personer i alderen 15-74 år.

Dækning

Den enkelte tælling omfatter et udsnit på 1,6 pct. af ovennævnte personkreds. Udvælgelsen er bestemt ud fra fødselsdatoer, der udskiftes ud fra et fastlagt mønster.

Datatyper

Oplysning om beskæftigelsesforhold mv. på tællingsdagen, oplysning om lediges arbejdssøgning, familieforhold, bopæl.

I enkelte undersøgelser er medtaget visse supplerende oplysninger, såsom (i 1979) uddannelsesoplysninger. Personnummer er identifikation.

Tidsmæssig dækning

Siden 1975.

Datagrundlag

1. Det centrale personregister.
2. Personligt besvarede postspørgeskemaer samt interviewbesvarede spørgeskemaer.
3. Befolkningsstatistikregistret.

Opdateringsfrekvens

Arligt.

12. Løn- og personalestatistikregistret for den offentlige sektor

Formål

Registret har til formål at give statistiske oplysninger om:

1. Antallet af ansatte i stat og kommune, herunder tilskudsområdet.
2. Løn- og beskæftigelsesforhold for ansatte i stat og kommune.

Definition af personkreds

Lønmodtagere ansat i staten og kommunerne, herunder tilskudsområdet.

Dækning

Total.

Datatyper

Registret indeholder for hver person oplysninger om ansættelsessted, ansættelsesforhold samt lønforhold. For kommunalt ansatte indhentes endvidere oplysninger om indenrigsministeriets autoriserede kontonummer, om vedkommende var ansat på et bestemt tidspunkt af året, samt om arbejdets omfang. Personnummeret er identifikation.

Tidsmæssig dækning

Registret dækker for staten årene siden 1971 og for kommunerne årene siden 1976. Indtil 1978 dækkede registret kun en 20% stikprøve.

Datagrundlag

Udtræk fra statslige og kommunale edb-lønssystemer samt spørgeskemaer fra de styrelser, kommuner mv., som ikke anvender edb-lønssystemer.

Opdateringsfrekvens

Arlig.

13. Arbejdsklassifikationsmodulet

Formål

Registret har til formål at give statistiske baggrundsoplysninger om personers beskæftigelsesforhold mv. til brug for de grene af personstatistikken, hvor sådanne baggrundsoplysninger er relevante.

Definition af personkreds

Alle personer med bopæl i Danmark samt andre personer der skal svare indkomst- eller formueskat til staten.

Dækning

Total.

Datatyper

Registret indeholder oplysninger om beskæftigelsesstatus og for erhvervsaktive tillige erhvervs- og stillingsoplysninger samt beskæftigelsesomfang. Grupperingerne foretages efter to forskellige principper: Efter væsentligste aktivitet i løbet af året og efter aktiviteten på en bestemt dato. Til yderligere karakteristik af beskæftigelsen indgår endvidere oplysning om indkomst og arbejdsgiver-virksomhedens størrelse. Identifikation er personnummer.

Tidsmæssig dækning

Fra 1976, dog også for 10 pct. stikprøve i 1975.

Datagrundlag

1. Det centrale skatteyderregister, persondel (CSR-P).
2. Det centrale oplysningsseddelregister (COR-S).
3. Sømandsskatteregisteret.
4. Det erhvervsstatistiske registersystem.
5. Arbejdsløshedsstatistikregisteret (lb.nr. 15).
6. Uddannelsesstatistikregisteret (lb.nr. 17).
7. Indkomststatistikregisteret (lb.nr. 6).
8. Arbejdspladsstatistikregisteret (lb.nr. 14).
9. Løn- og personalestatistikregisteret for den off. sektor (lb.nr. 12).

Opdateringsfrekvens

Årlig.

14. Arbejdspladsstatistikregistret

Formål

Registret skal danne grundlag for:

1. Statistik ang. beskæftigelsens geografiske og branchemæssige fordeling.
2. Pendlingsstatistik, dvs. opgørelser af befolkningen fordelt geografisk efter bopæl og arbejdssted.
3. Statistik ang. de lokale erhvervsenheders branche, personel mv.

Definition af personkreds

Registret indeholder oplysning om personer, der er eller har været ansat som lønmodtagere i Danmark.

Dækning

Total.

Datatyper

Registret indeholder oplysning om personernes arbejdsgiver, arbejdssted, beskæftigelsens omfang og lønindtægt. Personnummer er identifikation.

Tidsmæssig dækning

Fra og med året 1979.

Datagrundlag

1. Løn- og personalestatistikregistret for den offentlige sektor (lb.nr. 12).
2. Det centrale oplysningsseddelregister (COR-S).
3. Det erhvervsstatistiske registersystem.

Opdateringsfrekvens

Registret opdateres årligt.

15. Arbejdsløshedsstatistikregistret

Formål

At danne grundlag for udarbejdelse af detaljeret statistik til belysning af ledighedens struktur og udvikling set fra såvel en arbejdsmarkedsmæssig som social synsvinkel.

Definition af personkreds

Omfatter samtlige medlemmer af de statsanerkendte arbejdsløsheds-kasser samt andre personer, der tilmeldes et offentligt arbejdsformidlingskontor som arbejdssøgende.

Dækning

Total.

Datatyper

1. Oplysning om bopæl, ægteskabelig stilling, statsborgerskab, beskæftigelsesforhold, uddannelsesforhold, arbejdsløsheds-kas-setilhørsforhold og forsikringsforhold.
2. Ledighedsoplysninger, herunder om ledighedsårsag samt ledig-hedens varighed og omfang.

Tidsmæssig dækning

Fra 1. januar 1979.

Datagrundlag

Det centrale register for arbejdsmarkedsstatistik, CRAM.

Opdateringsfrekvens

Månedlig.

16. Selvangivelsesstatistikregistret

Formål

At danne grundlag for en detaljeret statistisk beskrivelse af de personlige indkomster, fradrag og formuer for forskellige befolkningsgrupper.

Definition af personkreds

Registret omfatter fuldt skattepligtige personer.

Dækning

En stikprøve udtaget blandt hovedpersoner (mænd og ugifte kvinder), hvorefter oplysninger for de gifte mænds hustruer indhentes samtidigt med oplysningerne for mændene (ca. 20.000-30.000 personer).

Datatyper

Registret indeholder detaljerede oplysninger om indkomster, fradrag og formuer samt hvert år (1971, 1973, 1975 og 1977) specielt detaljerede ejendomsoplysninger. Derudover indeholder registret oplysninger om køn, alder, civilstand, bopæl, arbejdsstilling samt fra 1974 oplysning om familieforhold. Personnummeret er identifikation.

Tidsmæssig dækning

Registret dækker perioden fra 1971.

Datagrundlag

1. De kommunale ligningsmyndigheder, som henter oplysninger om fradrag og formuespecifikationer fra de udfyldte selvangivelser.
2. Befolkningsstatistikregistret (lb.nr. 1).
3. Indkomststatistikregistret (lb.nr. 6)

Opdateringsfrekvens

Årlig.

17. Uddannelsesstatistikregistret

Formål

Registret har til formål at danne grundlag for en statistisk beskrivelse af elevernes uddannelsesplacering på et givet tidspunkt og af elevernes bevægelser gennem uddannelsessystemet i en given periode.

Definition af personkreds

Registret retter sig mod alle personer, der siden 1970 har været uddannelsessøgende i Danmark.

Dækning

Total.

Datatyper

Registret indeholder oplysninger om uddannelse, uddannelsesinstitution, påbegyndelses- og afgangstidspunkt, afsluttende eksamen mv., bopæl, civilstand og nationalitet. Personnummer er identifikation.

Delregistre

Der oparbejdes årligt 6 delregistre:

1. Skoler og kurser (8. kl. og opefter)
2. Frie skoler
3. Enkeltfagskurser
4. Lærlingeuddannelser
5. Visse erhvervsforberedende uddannelser
6. Videregående uddannelser

De vigtigste oplysninger samles pr. person i et syvende register, det integrerede elevregister.

Tidsmæssig dækning

Registret dækker siden 1973 uddannelsesforløbene på de fleste områder. Dog er universitetsuddannelserne først dækket fra 1975 og enkeltfagskurserne først fra 1978.

Endvidere dækkes de afsluttende eksaminer siden 1970.

Datagrundlag

Indrapportering fra skolerne.

Opdateringsfrekvens

Årlig.

18. Statistikregistret vedr. langvarigt uddannedes beskæftigelsesforhold

Formål

At give statistiske oplysninger om nyuddannedes forhold på arbejdsmarkedet.

Definition af personkreds

Personer med dansk statsborgerskab, der har taget en længerevarende uddannelse (ca. tre års uddannelse udover studentereksamens niveau) fra en dansk uddannelsesinstitution efter 1. oktober 1974.

Dækning

For uddannelser med over 1000 dimittender pr. år tages 50 pct. udsnit. For alle andre uddannelser er dækningen total.

Datatyper

Registret indeholder oplysninger om uddannelsesforhold (herunder eksamensresultat), om beskæftigelsesforhold siden dimissionstidspunktet, om arbejdssøgning og evt. videreuddannelse. Personnummer er identifikation.

Delregistre

1. 1-års undersøgelsen (omfatter personer, hvis dimissionstidspunkt ligger ca. et år tidligere end et tællingstidspunkt):
2. 3-års undersøgelsen (omfatter personer, hvis dimissionstidspunkt ligger ca. tre år tidligere end et tællingstidspunkt).

Tidsmæssig dækning

1. 1-års undersøgelsen omfatter personer dimitteret senere end 1. oktober 1976.
2. 3-års undersøgelsen omfatter personer dimitteret senere end 1. oktober 1974.

Datagrundlag

Personkredsen til en undersøgelse udtrækkes fra uddannelsesstatistikregistret (lb.nr. 17) med visse uddannelsesoplysninger. Adresseoplysninger fås fra CPR. Øvrige oplysninger indsamles ved spørgeskema.

Opdateringsfrekvens

Begge delregistre opdateres årligt.

19. Uddannelsesklassifikationsmodulet

Formål

Registret har til formål at give statistiske baggrundsoplysninger om befolkningens uddannelsesmæssige baggrund til brug inden for de personstatistikområder, hvor sådanne oplysninger er relevante.

Definition af personkreds

Registret retter sig mod den samlede befolkning med bopæl i Danmark.

Dækning

Total.

Datatyper

Afsluttende skole- og erhvervsuddannelse og evt. igangværende skole- og erhvervsuddannelse. Personnummer er identifikation.

Tidsmæssig dækning

Registret dækker årene fra 1979 og fremefter.

Datagrundlag

Uddannelsesstatistikregistret (lb.nr. 17) og folke- og boligællingsregistret (lb.nr. 3).

Opdateringsfrekvens

Årlig.

20. Bistandslovsstatistikregistret

Formål

At danne grundlag for statistik vedrørende ydelser efter bistandslovens kap. 8-11 samt enkelte andre ydelser.

Definition af personkreds

Familier, der modtager en af de pågældende ydelser.

Dækning

Total.

Datatyper

1. Baggrundsoplysninger: Personens bopæl, familieforhold, statsborgerskab, uddannelse. Sagens status pr. 1. januar og 31. december, ydelsens varighed samt de paragraffer, der er ydet bistand efter i året med angivelse af beløb.
2. Grundoplysninger for modtagere af kontanthjælp.
3. Supplerende oplysninger for modtagere af kontanthjælp: For et udsnit af personkredsen indeholder registersystemet nogle supplerende oplysninger, hvis disse personer har modtaget kontantbistand efter §§37, 40.1, 42.1, 43. Oplysningerne vedrører bistandens art (hvad ydelserne bruges til), forholdet til samtidighed mht. arbejdsløsheds- og sygedagpenge samt tilbagebetalingspligt.
4. Oplysninger om børn og unge: Forældremyndighed, ydelsesart, anbringelsessted, udskrivningssted, varighed.

Personnummer er identifikation.

Delregistre

1. Kontanthjælpsregistret: Grundoplysninger om kontanthjælp pr. familie.
2. Børneregistret: Grundoplysninger om bistand til børn og unge.
3. Supplementsregister: Grundoplysninger og supplerende oplysning for et udsnit af modtagerne af kontanthjælp.

Tidsmæssig dækning

Fra 1977.

Datagrundlag

Oplysningerne om bistand indhentes fra kommunerne dels på spørgeskema, dels i form af udtræk fra kommunernes edregistre. Der supple-

res med oplysninger fra befolkningsstatistikregistret (lb.nr. 1).

Endvidere indhentes enkelte oplysninger fra revaliderings- og pensionsnævnene.

Opdateringsfrekvens

Arligt.

Særlige bemærkninger

Personnumrene er krypteret.

21. Kriminalstatistikregistret

Formål

At give statistiske oplysninger om personer, der pålægges strafferetlige sanktioner.

Definition af personkreds

Registret omfatter personer, som pålægges en strafferetlig reaktion.

Dækning

Total.

Datatyper

Registret indeholder oplysning om kriminalitetsart, den pålagte reaktions art og længde/størrelse, forskellige baggrundsoplysninger som køn, alder, bopæl, uddannelses- og beskæftigelsesforhold, opvækstforhold samt oplysninger om, hvilke politi- og retsmyndigheder, den enkelte person/sag har været i berøring med. Personnummer er identifikation.

Tidsmæssig dækning

Registret går tilbage til 1972. Før 1979 omfatter registret ikke personnummer.

Datagrundlag

Udtræk af rigspolitichefens kriminalregister.

Opdateringsfrekvens

Årligt pr. 1. januar.

Bemærkninger

Personnumrene er krypteret.

22. Færdselsuhedsstatistikregistret

Formål

Registrets formål er at danne basis for færdselsuhedsstatistik, dvs. opgørelser af færdselsuheld og personskader som følge heraf med fordeling efter trafikale og demografiske kendetegn.

Definition af personkreds

Personer, som har været impliceret i færdselsuheld, som er kommet til politiets kendskab.

Dækning

Total i henhold til ovenstående definition. I forhold til alle færdselsuheld, altså også de som ikke kommer til politiets kendskab, er dækningen for de dræbte næsten total, for de alvorligt tilskadekomne ca. 70% og for de lettere tilskadekomne ca. 30%.

Datatyper

Registret indeholder i første led oplysninger om enkelte færdselsuheld med data for tid, sted, vejtype, uheldstype og omstændigheder i øvrigt. I andet led indgår det enkelte element i uheldet (køretøj, fodgænger). For motorkøretøjer findes registreringsnummer. I tredje led indgår personerne, identificeret ved personnummer og med oplysning om tilskadekomst og behandling, brug af sikkerhedsudstyr samt spirituspåvirkning mv.

Tidsmæssig dækning

Registret dækker årene fra og med 1967. De ældste besvarede år har ikke fuldt komplette personnummeroplysninger.

Datagrundlag

Dataindsamlingen udføres af politiet, der for hvert færdselsuheld udfylder et indberetningsskema med alle data i 4 kopier. Amtskommune eller kommune medvirker ved stedkodning af uheldene.

Opdateringsfrekvens

Månedlig.

23. Huslejepanelets register

Formål

At danne grundlag for de halvårslige huslejeundersøgelser, hvis resultater indgår ved beregning af indeks for boligudgiften i reguleringspristallet og i forbrugerprisindekset.

Definition af personkreds

Registret retter sig mod samtlige udlejningsboliger, beboet af private husstande.

Dækning

Repræsentativ stikprøve på ca. 5.000 lejligheder.

Datatyper

Registret indeholder lejers navn og stilling samt lejlighedens adresse, diverse data om lejligheden (alder, størrelse, udstyr, ejerform samt om den er underkastet huslejeregulering eller ej) og de oplyste lejebeløb.

Tidsmæssig dækning

Registret har oplysninger om huslejen siden oktober 1977.

Datagrundlag

Adressemateriale og lejlighedsoplysninger baseret på folketællingerne og tidligere interviews. Halvårsvis suppleres panelet med lejligheder i nybyggeri på grundlag af oplysninger fra kommunerne. Et mindre supplement af lejligheder beboet af pensionister i 1979 er baseret på et udtræk af CPR-registret.

Data om huslejen fås ved interviews med lejerne.

Opdateringsfrekvens

De variable oplysninger - huslejen - indhentes hvert halve år.

24. Udsnitsarkivet

Formål

At danne grundlag for at udnytte allerede indsamlet datamateriale i nye sammenhænge eller efter nye metoder, herunder også et beredskab for tidsforløbsanalyser.

Definition af personkreds

Registret retter sig mod den samlede befolkning med bopæl i Danmark.

Dækning

Udsnit af varierende størrelse udvalgt således, at kombination af oplysninger fra forskellige kilder og/eller tidspunkter er mulig.

Datatyper

Registret indeholder en række centrale statistiske personoplysninger, der kan henføres til en person. Personnummer er identifikation.

Delregistre

I registret indgår et delregister for hvert udsnit.

Tidsmæssig dækning

Det ældste materiale er fra 1969. Oplysningerne i udsnitsarkivet opbevares i 20 år.

Datagrundlag

Uddrag af en række af Danmarks Statistiks statistikregistre og i enkelte tilfælde direkte fra administrative registre, der i øvrigt ikke har været statistisk benyttede.

Opdateringsfrekvens

Nyt materiale indføres i registret på det tidspunkt, hvor der første gang er brug for det i en samkørsel, eller når det slettes i det oprindelige register.

Særlige bemærkninger

Statistiske uddrag af registret er til rådighed for forskning og planlægningsformål mv., uden for Danmarks Statistik, men alle kørsler udføres på Danmarks Statistiks edb-anlæg under medvirken af den ansvarlige for udsnitsarkivet.

Personnumrene er krypteret.

25. Medicinsk dødsfaldsstatistikregister for Grønland

Formål

Registret skal danne grundlag for statistik ang. dødsfald i Grønland.

Definition af personkreds

Registret retter sig mod den samlede befolkning med bopæl i Grønland. I registret optages oplysning om dødsfald i denne befolkning.

Dækning

Total.

Datatyper

Registret indeholder oplysning om:

1. De afdødes almindelige folkeregisterdata (bopæl, civilstand, nationalitet) på hændelsestidspunktet.
2. Oplysninger om dødstidspunkt og om medicinske forhold i forbindelse med dødsfaldet (dødsårsag mv.).

Tidsmæssig dækning

Fra 1975.

Datagrundlag

Oplysningerne tilvejebringes ved samkøring af oplysninger om døde personer fra CPR med oplysninger fra originale dødsattester.

Opdateringsfrekvens

Årlig.

26. Omnibusundersøgelsesregistret

Formål

Registret har til formål at tilvejebringe oplysninger til brug for videnskabelige og statistiske undersøgelser, der baseres på interview med mindre udsnit af den danske befolkning.

Definition af personkreds

Personer bosiddende i Danmark (excl. Grønland og Færøerne), som er fyldt 16 år.

Dækning

Stikprøver omfattende 2.500 eller 5.000 tilfældigt udvalgte personer.

Datatyper

1. Faste baggrundsoplysninger om interviewpersonernes køn, alder, familieforhold, beskæftigelsesforhold, skoleuddannelse, boligforhold, indkomstforhold, urbaniseringsgrad for bopæl.
2. Undersøgelsesoplysningerne dvs. personernes svar på undersøgelsesspørgsmålene. Som eksempler på emneområder, der har været dækket af en undersøgelse kan nævnes: forbrugerforventninger, udbredelsen af allergi, børnepasning.

Delregistre

1. Identifikationsregister, der indeholder nøgle mellem personnummer og løbenummer.
2. Undersøgelsesregister, der indeholder de omtalte baggrundsoplysninger og undersøgelsesoplysninger. Heri findes kun løbenr. som identifikation.

Tidsmæssig dækning

Omnibusundersøgelserne er gennemført regelmæssigt siden januar 1974.

Særlige bemærkninger

Danmarks Statistik er ansvarlig for registret i fællesskab med Socialforskningsinstituttet.

27. Forbrugsundersøgelserregistret

Formål

Registret har til formål at oparbejde et grundlag for statistik om befolkningens forbrugs-, indkomst- og opsparingsforhold i 1981, særligt til brug for reguleringspristallets vægtgrundlag.

Definition af personkreds

Registret omfatter et repræsentativt udsnit af alle landets husstande i 1981 med undtagelse af fælleshusstande.

Dækning

Undersøgelsen omfatter ca. 3.000 husstande.

Delregistre

1. Udsendelsesregister

Navn og adresse mv. for udtrukne husstande samt oplysning om deltagelse eller frafald.

2. Undersøgelsesregister

Oplysninger om forbrugsposter, indkomster mv.

Tidsmæssig dækning

Gælder 1981.

Særlige bemærkninger

Når forbrugsoplysningerne mv. er indsamlet og oparbejdet, slettes personidentifikationerne og registret ophører dermed.

Bilag 5: CPR's anvendelse i befolkningsstatistikken

I. Det befolkningsstatistiske system (BEF 78)

CPR har gennem adskillige år været den primære kilde for befolkningsstatistikken. Dette gælder især årene fra og med 1973, hvor der blev gennemført omfattende ændringer i statistiksystemet (BEF 73). I forbindelse med en større omlægning og forbedring af CPR i 1978 undergik befolkningsstatistiksystemet på ny en gennemgribende revision (BEF 78). I det følgende beskrives det nugældende system i hovedtræk, medens afsnit II omhandler forskellige problemer i forbindelse med registerudnyttelsen, herunder også specielle problemer i det tidligere system (BEF 73).

A. CPR-baserede opgørelser

Det centrale personregister opfattes i befolkningsstatistikken som en model af den danske befolkning:

Registret indeholder oplysninger om en objektgruppe bestående af samtlige personer med fast bopæl i Danmark eller Grønland. Oplysningerne om hver person dækker i alt væsentligt de befolkningsstatistiske behov, og datakvaliteten er generelt set høj. Endvidere gælder det, at CPR ajourføres løbende, og normalt sker det med en meget lille forsinkelse i forhold til hændelsestidspunktet.

CPR ajourføres dagligt (ugens fire første dage) på grundlag af ændringsmeddelelser fra kommunerne. I forbindelse med ajourføringen af personregistret kan CPR's brugere modtage tilstandsudtræk, dvs. udtræk af det opdaterede registers indhold på det aktuelle tidspunkt. Desuden kan der leveres ændringsudtræk, som for de personer, hvis data er ændret, angiver ændringens art samt dataindholdet før og efter ændringen.

Analogt hermed kan der rekvireres hhv. tilstands- og ændringsudtræk fra vejregistret. Vejregistret indeholder oplysning om samtlige veje/vejstykkers placering, dels i visse faste områdeinddelinger (postdistrikter, sogne osv.), dels i specielle, af kommunen fastlagte, inddelinger såsom skoledistrikter og planområder mv.

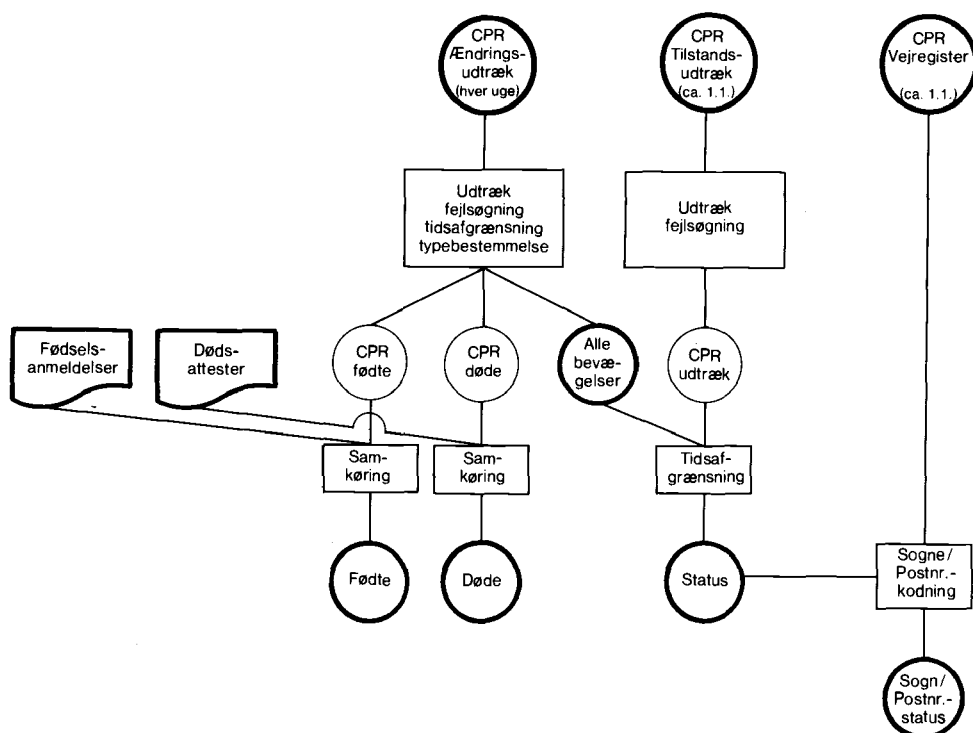
Til brug for den løbende befolkningsstatistik modtager Danmarks Statistik ved hvert årsskifte en total kopi af CPR's personregister og vejregister. Endvidere modtages året rundt ugentlige ændringsudtræk, idet CPR opsamler de i ugens løb foretagne

ændringer.

Det CPR-baserede statistiksystem er skitseret i fig. 1. Ændringsudtrækkene danner grundlag for den kvartalsvise statistik over befolkningens bevægelser i de enkelte kommuner, og indgår desuden som grundlag for årsstatistikken. Der foretages udtræk og fejlsøgning af oplysningerne fra CPR, hvorefter der tidsafgrænses, hvilket sker på den måde, at alle ændringer, som ajourføres i CPR med en forsinkelse på 40 dage eller derover efter en periodes afslutning, overføres til den følgende periode. Hvis forsinkelsen derimod er under 40 dage, bliver ændringen medtaget i den statistikperiode, hvor den fandt sted. Ved således at vente 40 dage på forsinkede indberetninger opnås, at alle oplysninger, som ikke er ekstraordinært forsinkede, bliver henført til det rigtige hændelsestidspunkt. (Den gennemsnitlige forsinkelse - fra hændelsestidspunktet til registreringen i CPR - er ca. 2 uger).

Udover tidsafgrænsningen foretages der en præcis typebestemmelse af ændringerne. Dette indebærer, at det for hver ændringsrecord må fastslås, hvilken type hændelse der er tale om: fødsel, dødsfald, vandring, vielse osv. De enkelte ændringsmeddelelser

Fig. 1: Dannelse af befolkningsstatistikens grundmaterialer.



fra CPR er ganske vist på forhånd typebestemt, men det er nødvendigt at foretage en række justeringer mv., således at typeangivelsen bliver i overensstemmelse med befolkningsstatistikens behov. Der dannes herigennem tre magnetbåndsfiler, hvoraf den ene indeholder samtlige transaktioner, medens de to andre har særskilte oplysninger vedrørende hhv. fødte og døde. Filen, som indeholder oplysning om samtlige bevægelser, anvendes direkte til den kvartalsvise statistik over befolkningens bevægelser på kommuneniveau og anvendes desuden i den årlige befolkningsstatistik. Den videre behandling af fødte- og dødefilerne omtales nedenfor i afsnit B.

Statusopgørelserne over befolkningens størrelse og sammensætning på et givet tidspunkt foretages med udgangspunkt i en to-talkopi af CPR's personregister, som modtages fra CPR hvert år omkring den 1.1. Materialet omfatter ikke alene de i befolkningen faktisk tilstedeværende personer, men også såkaldte inaktive personer (dvs. døde, udvandrede og forsvundne). Desuden er der for de enkelte personer en række oplysninger, som ikke kan bruges i befolkningsstatistikken. Der foretages derfor først udtræk og fejlsøgning af relevante oplysninger for alle aktive personer. Dernæst skal den udtrukne population og de tilhørende oplysninger tidsafgrænses, idet der må tages hensyn til, at indberetningerne til CPR som nævnt sker med en vis forsinkelse. Ved hjælp af bevægelsesoplysningerne frem til 40 dage efter den 1. januar korrigeres CPR-udtrækket, således at det bringes i bedst mulig overensstemmelse med de faktiske befolkningsforhold pr. 1.1. Eksempelvis vil en fødsel, der er indtruffet den 30. december, men ikke er ajourført i den modtagne CPR-kopi (kopien kan fx være fremstillet den 5. januar det følgende år, medens fødslen fx først er indberettet (og ajourført) den 10. januar), i tidsafgrænsningen give anledning til, at den fødte med tilhørende personoplysninger overføres til den dannede status pr. 1.1. Statusfilen kan i første omgang benyttes til opgørelser på kommuneniveau, idet kommunekoden er indeholdt i den enkelte persons adressekode.

Opgørelser på mindre områder end kommuner kræver oplysning om adressernes placering i forhold til de pågældende områdeinddelinger. Vejregistret holder som nævnt rede på veje/vejstykkers beliggenhed i relation til visse faste områdeinddelinger, bl.a. sogne og postdistrikter, og ved at overføre disse oplysninger til personerne i status, udarbejdes den årlige sogne- og postnummer-status. På samme måde kan foretages subkommunale opgørelser af

befolkningens bevægelser. Metoden anvendes i servicesystemet KÅS til fremstilling af status- og bevægelsesstatistik fordelt på vilkårlige geografiske områder.

B. Anvendelse af andre kilder

Indtil nu er kun omtalt den rent CPR-baserede befolkningsstatistik. Til brug for de endelige opgørelser af fødte og døde suppleres CPR-dataene imidlertid med oplysninger fra andre kilder. Dette sker ved samkøring af CPR-oplysningerne med attestmaterialer modtaget fra Sundhedsstyrelsen (SHS). Disse samkøringer er beskrevet i bilag 2, eksempel 8 og 9 og skal derfor kun berøres ganske kort her. Samkøringen af CPR-materialet og SHS-materialet (for fødte fødselsanmeldelser, for døde dødsattester) foretages med henblik på at koble medicinske oplysninger om fødslerne og dødsfaldene sammen med de demografiske CPR-data. Sammenkoblingerne sker for dødsfaldene ved hjælp af afdødes personnummer, som findes i begge materialer, og for fødslerne ved hjælp af moderens personnummer. Sammenkoblingen sker med meget stor præcision, hvilket bl.a. skyldes, at de indgående datamaterialer i vidt omfang har samme oprindelse.

Efter overgangen til BEF-78 er kun to områder indenfor befolkningsstatistikken uden direkte forbindelse med CPR. Det drejer sig dels om abortstatistikken, hvor kildematerialet er anmeldelser fra sygehusene, og dels om statistikken vedrørende papirløse samlivsforhold, hvor oplysningerne indsamles gennem de udnitsbaserede omnibusundersøgelser.

2. Problemer i forbindelse med registerudnyttelsen

A. Indberetningssystemet (ex.: Vielser/skilsmisser)

Som en konsekvens af, at ændringsindberetningerne i det tidligere CPR-system (før september 78) ikke kom som direkte meddelelser om hændelser af bestemte typer, men derimod som indberetninger om ændringer i de enkelte personers dataindhold, havde man i BEF 73-systemet visse problemer med at bestemme antallet af hændelser. Et sådant problem er omtalt i bilag 2, eksempel 6, som vedrører samkøring af viede personer til et antal vielser. Et helt analogt problem gjaldt opgørelsen af antal skilsmisser. Problemet opstod, fordi det ikke i alle tilfælde var muligt at finde en "makker" til den viede eller skilte person. Dette kunne i nogle tilfælde være helt korrekt, enten fordi makkeren var bosat i udlandet, eller fordi oplysningen om hændelsen for ægtefællen var

forsinket. Kun i meget få tilfælde var der tale om egentlige fejlindberetninger til CPR.

I det nye CPR-system (fra sept. 78), er problemet principielt ryddet af vejen, idet kommunerne nu indrapporterer veldefinerede hændelser til CPR i stedet for de tidligere mere uspecificerede dataændringer. Det vil fx sige, at den enkelte skilsmisse altid indberettes til CPR som en hændelse med samtidig oplysning om både manden og kvinden.

B. Korrektioner og annullationer

Korrektioner kan defineres som dataændringer, der ikke modsvares af reelle hændelser. Fx kan en kommune foretage ændringer i vejcodesystemet eller ændringer i husnummereringen. Et andet eksempel er rettelse af en fejlagtigt registreret civilstandskode (fx fra ugift til gift). Sådanne korrektioner skal så vidt muligt holdes udenfor de statistiske opgørelser af hhv. flytninger og vielser.

Annullationer kan opfattes som specialtilfælde af korrektioner, idet en annullation har til formål at eliminere virkningen af en tidligere indberettet hændelse, som rent faktisk ikke har fundet sted. Eksempelvis forekommer det, at en person anmelder udvandring til folkeregistret, men af en eller anden grund aldrig gennemfører udvandringen. Når folkeregistret senere, ofte ved et tilfælde, opdager, at personen ikke er udvandret, skal folkeregistret annullere den tidligere indberettede udvandring.

Under det tidligere CPR-system var det ikke muligt præcist at udskille korrektioner og annullationer fra reelle hændelser. Særligt generende var de nævnte administrativt foretagne ændringer af vejkode og husnumre, idet disse ændringer - som forekom i betydeligt antal - kun i begrænset omfang kunne adskilles fra flytninger indenfor kommuner. Som en konsekvens heraf undlod man stort set at publicere oplysninger vedrørende subkommunale flytninger.

I det nuværende CPR-system, hvor de enkelte ændringsindberetninger principielt er veldefinerede, kan korrektioner og annullationer behandles særskilt, hvilket generelt har medført en kvalitetsforbedring. Der er endvidere blevet basis for at fremstille en komplet flyttestatistik omfattende samtlige flytninger - også subkommunale.

I BEF 78 indgår korrektioner og annullationer ikke i de lø-

bende opgørelser af befolkningens bevægelser (hændelsesstatistikken), hvorimod de medregnes i forbindelse med tidsafgrænsningen af befolkningsstatus. Principielt burde man løbende korrigere hændelsesstatistikken for indberettede annullationer, men dette undlades, først og fremmest på grund af det generelt meget begrænsede antal annullationer.

C. Hændelsesdatoen

Hændelsesdatoen er naturligvis af vital betydning for statistikken. Der konstateres imidlertid et vist antal fejl i ændringsmeddelelsernes hændelsesdato. Fejlene kan i visse tilfælde afsløres og rettes, fx hvis der forekommer umulige kombinationer af fx fødselsdato og hændelsesdato eller alder og hændelsesdato. Men det er klart, at der forekommer flere fejl, som blot ikke kan afsløres, idet en fejl jo ikke behøver at resultere i umulige eller usandsynlige datakombinationer.

I CPR er problemet dog meget begrænset, da registret som nævnt normalt opdateres hurtigt og præcist.

D. Kommunikation med registermyndigheden

Det er vigtigt for brugeren at holde sig i kontakt med registermyndigheden, således at fx ændringer i registerdriften - og registerindholdet - ikke kommer som ubehagelige overraskelser. Gennem et permanent CPR-kontaktudvalg opretholdes derfor en løbende forbindelse mellem Danmarks Statistik og Sekretariatet for personregistrering (Sfp), som bestyrer registret.

Alligevel kan der ske misforståelser og fejltagelser som følge af manglende informationer, hvilket naturligt nok især forekom i overgangsfasen til befolkningsstatistikens intensive udnyttelse af CPR. Men faren er selvfølgelig altid til stede. Et eksempel: I foråret 1977 opfordrede Sfp kommunerne til at dørunummerere lejligheder i de tilfælde, hvor der befandt sig mere end tre lejligheder pr. etage. Dørunumrene skulle erstatte sideangivelserne, og ændringen blev foretaget for at opnå overensstemmelse med registreringer i bygnings- og boligregistret. Ændringen blev imidlertid foretaget på en måde, som Danmarks Statistik ikke var vidende om: For de lejligheder, der blev dørunummereret, placeredes dørunummeret i et særligt felt i adressen, "værelsesnr.", og samtidig blankstilledes det felt, som tidligere indeholdt sideangivelsen. Denne ændring medførte, at opgørelser af antal "adressehusstande" - dvs. af antal beboede adresser - ville give et ukorrekt resul-

tat, hvis man ikke tog hensyn til feltet "værelsesnummer". I servicesystemet indgår en tabel vedrørende antal adressehusstande, men den omtalte ændring betød, at man i dette års KÅS-statistik for nogle kommuner pludselig fik et drastisk fald i antal adressehusstande, idet lejlighederne på de etager, som var dørunnummeret, blev opfattet som en enkelt adresse.

Bilag 6: Klassifikationsmoduler i Danmarks Statistik

1. Befolkning/datamodul

Formålet med dette modul er at muliggøre afgrænsning af befolkningen til at være de personer, som havde bopæl i Danmark på et vilkårligt tidspunkt.

Grundlaget er statusuddrag på givne tidspunkter samt uddrag af løbende ændringer mellem disse fra det centrale personregister (CPR). Statusudraget indeholder blandt andet personnummer for hver person, som havde bopæl i landet på statustidspunktet. Ændringerne er ligeledes identificeret og indeholder oplysninger om ændringens art (fødsel, død, udvandring mm.) samt dato for ændringen. En given status kan således føres frem eller tilbage på grundlag af nøjagtige tidsangivelser for ændringerne.

2. Områdekodningsmodul

Personstatistikregistrene indeholder almindeligvis oplysning om bopælskommune til anvendelse ved geografiske opdelinger. Områdekodningsmodulet gør det muligt at operere med andre områdeinddelinger helt efter ønske, fx sogne, postdistrikter og skole-distrikter.

Modulet kan behandle såvel bopælsadresser som arbejdsstedsadresser, således at områdeinddelinger kan anvendes ved opgørelser fordelt efter bopæl og fordelt efter arbejdssted.

3. Familie/husstandsmodulet

Modulet grupperer personerne i familieenheder. Grundlaget er henvisninger fra børn til forældre og mellem ægtefæller samt bopælskoden, som alle fremgår af det under afsnit 1. omtalte statusuddrag. Der opereres med to familiebegreber. Det snævre begreb omfatter familietyperne enlige og ægtepar (ægtefæller med samme bopæl), begge typer med eventuelle børn under 18 år på samme bopæl. Det bredere begreb omfatter familietyperne enlige og ægtepar med eventuelle børn under 26 år på samme adresse samt ugifte samboende med fælles børn (under 26 år).

Modulet kan endvidere gruppere personerne i husstande, hvor en husstand defineres som en gruppe personer, der bor på samme adresse.

Familie/husstandsmodulet kan bruges til at knytte familieoplysninger til en given personkreds.

4. Uddannelsesklassifikationsmodulet (UKM)

UKM har samme karakter som arbejdsklassifikationsmodulet. UKM muliggør klassificering af enkeltpersoner efter højeste afsluttede almenuddannelse, evt. afsluttet erhvervsuddannelse og evt. igangværende uddannelse.

Oplysningerne indhentes fra uddannelsesinstitutioner, og der indgår endvidere oplysninger fra folketællingen 1970 specielt til dækning af de ældre årgange.

5. Arbejdsstedsmodul

Det såkaldte arbejdspladsstatistikregister indeholder en arbejdsstedsidentifikation for alle lønmodtagere. Erhvervsregistret indeholder tilsvarende identifikation sammen med personnummer for indehavere. De to registre udgør således tilsammen et modul, gennem hvilket personer kan grupperes efter arbejdssted, geografisk og med hensyn til branche. Denne mulighed skal bl.a. bruges til løbende pendlingsstatistik.

6. Arbejdsklassifikationsmodulet (AKM)

Modulet klassificerer den enkelte person efter beskæftigelsesstatus, beskæftigelsens art samt arten af det arbejdssted, hvor den pågældende er beskæftiget. Der klassificeres efter væsentligste beskæftigelse i løbet af året og efter beskæftigelsen på en given dag i året.

Bilag 7: Arbejdsklassifikationsmodulet (AKM)

1. Formål

AKM, som er dannet årligt siden 1976, indeholder oplysninger, der klassificerer den enkelte person efter beskæftigelsesstatus, beskæftigelsens art, arten af den virksomhed, hvori den pågældende er beskæftiget samt til en vis grad beskæftigelsens omfang ved denne virksomhed. Endvidere indgår enkelte andre beskæftigelsesrelevante klassifikationer. Oplysningerne er identificeret ved personnummer, således at de kan overføres til anvendelse i eller sammen med forskellige personstatistiske registre.

Klassificeringen foretages efter to principper, idet der klassificeres efter dels væsentligste beskæftigelse i løbet af året og dels beskæftigelsen på en given dag i året.

2. Grundoplysningerne

AKM er dannet på grundlag af oplysninger fra en række administrative registre, som dog i visse tilfælde har gjort vejen om ad andre statistikregistre i Danmarks Statistik. Der findes i disse registre forskellige oplysninger, som kan bidrage til en beskrivelse af beskæftigelsen for en given person, men ingen af disse er isoleret set tilstrækkelig til at foretage klassificering efter de i afsnit 1 omtalte begreber. Oplysningerne fra forskellige kilder kædes sammen for den enkelte person, og der foretages en afvejning efter et bestemt regelsæt.

Fra ni forskellige registersystemer anvendes følgende grundoplysninger for den enkelte person:

- a. overskud af selvstændig erhvervsvirksomhed
- b. indkomst som medhjælpende ægtefælle
- c. løn (pr. arbejdsgiver)
- d. pension
- e. ansættelsesforhold identificeret ved arbejdsgivernummer
- f. beskæftigelsesperioder
- g. bidrag til pensionsfond (ATP)
- h. regnskabspligt
- i. stillingsbetegnelse (fra centralt personregister)
- j. stillingsoplysninger (offentligt ansatte)
- k. stillingsoplysninger (sømænd)
- l. kode for arbejdsledelseskasse
- m. arbejdsledelsesperioder

- n. uddannelsesoplysninger
- o. brancheoplysninger (for hvert ejer- og/eller ansættelsesforhold)

3. Klassifikationer

3.1. Beskæftigelsesstatus.

Med dette begreb deles befolkningen i hovedgrupperne:

- selvstændige
- medhjælpende ægtefælle
- lønmodtagere
- pensionister
- øvrige ude af erhverv

I denne gruppering indgår, som det ses, den primære sondring: i erhverv / ude af erhverv.

Ved klassifikation efter årsprincippet tages udgangspunkt i, hvilke personer, der i det erhvervsstatistiske registersystem står anført som virksomhedsindehaver, der har været aktiv i løbet af det omhandlede år. For disse personer foretages en afvejning af forskellige indkomstarter mv. (fra indkomstskattesystemet), og de personer, for hvem det gælder, at indkomstsammensætningen ikke taler væsentligt imod selvstændig erhvervsvirksomhed som vigtigste indkomstkilde, klassificeres som selvstændig.

Det erhvervsstatistiske registersystem omfatter ikke alle selvstændige erhvervsdrivende, kun de, som tilbageholder kildeskat (arbejdsgivere), eller som er momspligtige, indgår. Den mindre gruppe selvstændige, som falder uden for denne kreds, afgrænses på grundlag af indkomstoplysninger samt oplysning om eventuel regnskabspligt.

De øvrige grupper klassificeres efter største indkomsts princip, idet indkomstarterne virksomhedsoverskud, indkomst som medhjælpende ægtefælle, løn og pension sammenlignes. For personer med begrænset lønindkomst er der en undergrænse for, hvornår de klassificeres som erhvervsaktive. Ved denne afgrænsning indregnes også lønkompenenserende ydelser som arbejdsløsheds- og sygedagpenge i den samlede løn.

Klassifikation efter datoprincippet foretages på grundlag af oplysninger i arbejdsgivernes indberetning til skattemyndighederne (oplysningssedlerne), som blandt andet indeholder

ansættelsespersoner for de enkelte ansatte, det erhvervsstatistiske registersystems oplysninger om start- og stopdatoer for de enkelte virksomheder samt oplysninger om arbejdsløshedsperioder fra arbejdsløshedsstatistikregistret. Gennem anvendelse af periodeangivelserne udtrækkes de oplysninger, som er relevante i forhold til opgørelsesdatoen, og der foretages herefter en vis begrænset indkomstafvejning samt en prioritering af eventuelle flere muligheder, hvor der tages højde for, hvilke af de foreliggende grundoplysninger, der kan tillægges størst vægt.

3.2. Stillingsklassificering.

For den persongruppe, der i beskæftigelsesstatus er klassificeret som lønmodtager, foretages en mere specifik klassifikation efter beskæftigelsens art.

Der foreligger stillingsoplysninger fra tre kilder, og der anvendes støtteoplysninger fra yderligere to kilder. Kun stillingsoplysningerne fra det centrale personregister dækker hele den voksne befolkning. De to øvrige kilder omfatter delpopulationerne offentligt ansatte og sømandsskattepligtige. Støtteoplysningerne er dels en eventuel kode for, hvilken arbejdsløshedskasse en given person er medlem af (en vis faglig afgrænsning af kasserne giver nogle retningslinier for, hvilke typer stillingsbetegnelser der kan være tale om), og dels en kode for, i hvilken branche vedkommende er beskæftiget. De to støtteoplysninger benyttes, hvor de primære stillingsoplysninger er uklare, og til en vis opdeling af personer, der ellers ville være klassificeret som uoplyst.

Med henblik på behandlingen af disse stillingsoplysninger er der i tilknytning til AKM oprettet omsætningstabeller, der i første trin omformer de uensartede primæroplysninger til standardkoder. I andet trin vil der ske en udvælgelse blandt de foreliggende muligheder, og her vil støtteoplysningerne indgå. Hele omsætningsforløbet vil resultere i, at hver person tildeles én standardkode som udtryk for den stillingsklassificering, der er udledt af oplysningerne fra de forskellige kilder.

I sin hidtidige form har AKM's stillingsklassifikation alene været baseret på stillingsoplysninger fra det centrale personregister, og det udvidede grundlag vil først indgå ved

dannelsen af AKM-80 i slutningen af 1981. Den endelige udformning af omsætningssystemet er derfor ikke fastlagt i detaljer. Det kan fx blive aktuelt at supplere med uddannelsesoplysninger til støtte for stillingsklassificeringen.

Primæroplysningerne vedrørende stilling refererer til ultimo det pågældende år. Der benyttes samme grundlag i den årsbaserede og den datobaserede klassifikation. (Beskæftigelsesstatus kan dog være forskellig efter de to principper, således at lønmodtagergruppen som helhed er forskelligt afgrænset).

Som en slags stillingsklassificering gennemføres i visse AKM-versioner en opdeling af gruppen "øvrige ude af erhverv", idet oplysninger om igangværende uddannelse indhentes fra uddannelsesklassifikationsmodulet.

3.3. Erhvervsklassificering.

Arbejdsgiverindberetningerne til indkomstskattesystemet vedrørende udbetalt løn, ansættelsesperiode mm. (de tidligere omtalte oplysningssedler) identificerer såvel arbejdsgiver som lønmodtager i et givet ansættelsesforhold. Gennem et særligt projekt (arbejdspladsprojektet) er der yderligere skabt korrespondance mellem den enkelte lønmodtager og det enkelte arbejdssted blandt eventuelt flere inden for samme arbejdsgiver. Det er således muligt at overføre den branchekode, som gælder for et givet arbejdssted til hver enkelt af de ansatte.

Ved erhvervsklassifikation efter års-princippet identificeres det arbejdssted, hvor lønmodtageren har haft den største indkomst, og branchekoden for dette overføres via de nævnte identifikationer til den pågældende lønmodtagers AKM-record.

På tilsvarende måde foretages erhvervsklassifikationen efter datoprincippet, ud fra branchen på det arbejdssted, hvor lønmodtageren var beskæftiget den pågældende dato.

Arbejdspladsprojektet, der befinder sig i indkøringsfasen, har endnu ikke afgivet oplysninger til AKM. Det vil ske på et senere tidspunkt for 1979-versionen og løbende ved fremtidige versioner.

For selvstændige overføres branchekoden direkte fra det

erhvervsstatistiske registersystem, og medhjælpende ægtefæller klassificeres via henvisning til ægtefællen. Selvstændige, der ikke kan klassificeres via registersystemet, fordi de hverken er arbejdsgiver eller momspligtig, klassificeres på grundlag af stillingsbetegnelsen fra personregistret.

3.4. Øvrige klassifikationer.

Indbetalinger af bidrag til arbejdsmarkedets tillægspensionsfond (ATP) administreres af arbejdsgiverne, og det samlede indbetalte bidrag oplyses for hver ansat på den tidligere omtalte oplysningsseddel. Bidragsstørrelsen af afhængig af antal udførte arbejdstimer på en sådan måde, at det kan anvendes som et tilnærmet udtryk for beskæftigelsens omfang for den enkelte lønmodtager. Der kan dog ikke alene udfra ATP-bidraget sondres mellem deltidbeskæftigelse og periodisk beskæftigelse, jf. afsnit 5. Summeret pr. arbejdssted anvendes ATP-bidraget som udtryk for arbejdsstedets størrelse.

Andre klassifikationer er dannet som faste standarder ved kombination og/eller sammentrækning af én eller flere af de ovennævnte klassifikationer.

4. Hittidig anvendelse af AKM

Modulet er som nævnt etableret med henblik på en bred anvendelse inden for personstatistikken. Hidtil har det været benyttet i forbindelse med registerfolketællingen i 1976 og i den årlige statistik vedrørende personlige indkomster og formuer. Endvidere har det sammen med befolkningsstatistikregistret og visse andre moduler været grundlag for gennemførelsen af et særligt serviceprogram (KAS INDKOMST), der tilbydes kommuner, hvori indgår fordelinger efter demografiske samt indkomst- og beskæftigelsesmæssige kendetegn. Endelig er modulet blevet anvendt ved mindre serviceopgaver, hvor der har været behov for opdelinger efter beskæftigelsesforhold.

5. Svagheder i AKM

Som nævnt er AKM under udbygning, så det er næppe alle problemer i den beskrevne version, der endnu er kendt. Det hidtidige arbejde med de foreliggende versioner har imidlertid afsløret, at problemerne knytter sig til især tre forhold, nemlig grunddatas egnethed, grunddatas pålidelighed og den manglende konsistens mellem visse oplysninger.

Egnethed drejer sig om, i hvilken grad grunddata begrebsmæssigt er i overensstemmelse med den anvendelse, der gøres af dem i klassifikationssystemet.

Et eksempel på begrænset egnethed findes i det ovenfor omtalte ATP-bidrag. Det udtrykker noget om beskæftigelsesomfanget, men det er bundet til regler om antal udførte arbejdstimer, således at der ikke betales bidrag ved mindre end 10 timers beskæftigelse pr. uge, 1/3 bidrag ved 10-19 timers beskæftigelse, 2/3 bidrag ved 20-29 timers beskæftigelse og fuldt bidrag ved 30 eller flere timers beskæftigelse. På den baggrund er det vanskeligt at omsætte ATP-bidraget til arbejdstimer for den enkelte, hvortil kommer, at man ikke kan aflæse, om det drejer sig om deltids- eller periodisk beskæftigelse, hvis ATP-bidraget er begrænset.

Et andet og mere tungtvejende eksempel er stillingsbetegnelserne fra personregistret. De er ikke indholdsmæssigt bundet af administrative regler, og den enkelte borger kan frit bestemme, hvad der skal anføres som stillingsbetegnelse. Det betyder, at stillingsbetegnelserne kan variere indholdsmæssigt fra at være egentligt beskæftigelsesorienteret over uddannelsesorientering til at være intetsigende titler. Når betegnelserne fjerner sig fra det førstnævnte, opstår klassifikationsproblemerne. Ved udvidelsen af klassifikationsgrundlaget vil det dog formentlig vise sig, at en del betegnelser, der isoleret set er indholdsløse, er mere egnede, når de ses i sammenhæng med andre oplysninger.

De anførte eksempler viser, hvor de større problemer vedrørende egnethed ligger. For de øvrige grunddata er egnetheden variende men stort set rimelig god.

Pålideligheden af oplysninger, der indsamles som led i en administrativ proces, er stærkt afhængig af, hvilke konsekvenser oplysningerne kan have i den administrative proces. Når det fx gælder de indkomstoplysninger og andet, som indgår i skatteligningen, lægges der stor vægt på, at de afgivne oplysninger er korrekte. Pålideligheden er i sådanne tilfælde stor, selv om der kan forekomme uafsluttede sager med deraf følgende ufuldstændige oplysninger.

Det modsatte yderpunkt er oplysninger, som ikke har nogen umiddelbar administrativ betydning, eller som ikke på anden måde kontrolleres og følges op, hvor der måtte være fejl eller mangler. Stillingsbetegnelserne indberettet fra borgerne via skattesystemet til personregistret er et eksempel på sådanne oplysninger. Konsekvensen er her,

at et stort antal personer, der efter de øvrige foreliggende oplysninger skulle have en stillingsbetegnelse, mangler en sådan, og man må formode, at der findes fejlagtige betegnelser i et vist antal tilfælde.

Konsistensproblemet opstår som følge af, at tidsreferencen for de anvendte grunddata ikke er ens. Det kommer tydeligt til udtryk ved klassifikation efter års-princippet, hvor stillingsoplysninger, der refererer til slutningen af året, kan optræde i sammenhæng med fx brancheoplysninger, der har været gældende i begyndelsen af året. Det er imidlertid et problem, der ikke rigtigt lader sig løse, idet man ikke kan udsøge en betegnelse for den væsentligste arbejdsstilling i løbet af året, svarende til brancheklassifikationen.

Bilag 8: Milepæle i registerstatistikken siden 1968

Danmarks Statistiks virksomhed i almindelighed og de enkelte statistikområder og omlægninger heri omtales udførligt i den af Danmarks Statistik årligt udgivne "Beretning og Arbejdsplan". I det følgende omtales nogle hovedpunkter i udviklingen, specielt for så vidt angår anvendelsen af edbteknik og registermetoder:

1. Med virkning fra 1968 etableredes en statistik vedrørende omsætning mv. i samtlige erhverv, baseret på oplysninger fra momsadministrationen kombineret med data fra Danmarks Statistiks erhvervsregister.

Fra 1971 etableredes en tilsvarende statistik vedrørende erhvervenes beskæftigelse og fra 1974 tillige om udbetalt lønsum baseret på arbejdsgiverens indberetninger til ATP-ordningen henholdsvis skattemyndighederne.

Efter etableringen af disse opgørelser er udarbejdelse af generelle "erhvervstællinger", der blev udarbejdet for 1925, 1935, 1948 og 1958, ophørt.

2. Pr. 1.5.1970 udarbejdedes første gang en totalopgørelse af befolkningen i de enkelte kommuner med fordeling på køn, alder og ægteskabelig stilling. I perioden 1970-73 gennemførtes en omlægning af den øvrige befolkningsstatistik til CPR-basis, således at manuelle indberetninger fra kommunerne kunne ophøre.
3. I 1971 anskaffede Danmarks Statistik eget edb-anlæg.
4. Fra 1970 omlagdes indkomststatistikken i forbindelse med kilde-skattens indførelse til registerbasis. Fra 1976 rekonstrueredes statistikken væsentligt, bl.a. ved at indkomstoplysningerne samkørtes med oplysninger fra et arbejdsklassifikationssystem, som også anvendes til andre formål, jf. punkt 7.
5. I 1971 etablerede Danmarks Statistik et "udsnittsarkiv" til brug for udviklingen af registerstatistiske metoder, herunder vedrørende "samkøring", og for udarbejdelse af analyser og ad hoc opgørelser.
6. Efter beslutning i 1971 overtog Danmarks Statistik i 1973 fra undervisningsministeriet statistikken vedrørende elever og studerende på skoler og fra undervisningsministeriet. De individuelle oplysninger, som undervisningsinstitutionerne indberetter, indgår i et uddannelsesstatistisk register, således at institutionernes løbende indberetninger kun skal omfatte ændringer i forhold til

året i forvejen.

7. I 1974 besluttede økonomiministeren, at der ikke skulle gennemføres "almindelig folketælling" i folketællingsterminen 1975/76, jf. § 7 i Lov om Danmarks Statistik. I stedet gennemførtes for 1976 en "registerfolketælling" på basis af CPR-oplysninger kombineret med en række forskellige andre registeroplysninger vedrørende erhvervsforhold mv.
8. Fra 1977 etableredes en særlig bistandslovsstatistik baseret på oplysninger indhentet på skemaer fra kommunerne, som skulle kombineres med oplysninger fra det kommunale "økonomisystem" og oplysninger, som Danmarks Statistik i forvejen var i besiddelse af, fx familieoplysninger fra CPR.
9. Pr. 1.4.1977 gennemførtes den første bygnings- og boligopgørelse på basis af det nyoprettede Bygnings- og boligregister. Pr. 1. januar 1980 gennemførtes den første registerbaserede bolig-tælling, som kombinerer oplysninger om boligerne med oplysninger om beboerne.
10. I 1979 besluttedes at der skal oprettes et "Arbejdspladsstatistikregister" baseret på eksisterende registeroplysninger og en begrænset supplerende dataindsamling med bistand af skattemyndighederne og arbejdsgiverne. Dette indebærer bl.a., at totale dataindsamlinger fra befolkningen i form af traditionelle folketællinger eller lign. definitivt kan ophøre.
11. Pr. 1.1.1981 gennemførtes den første egentlige folke- og bolig-tælling på registerbasis.

Bilag 9: Det erhvervsstatistiske registersystem

1. Baggrund mv.

Med henblik dels på indhentning af oplysninger til statistisk brug, dels på bearbejdning og samkøring af sådanne oplysninger, har Danmarks Statistik opbygget et edb-registersystem. Registersystemet betegnes det erhvervsstatistiske registersystem og omfatter samtlige de virksomheder, der indgår i erhvervsstatistikken, og det indeholder en række forskellige oplysninger om virksomhederne. Oplysningerne indhentes med hjemmel i loven om Danmarks Statistik fra såvel administrative registre som fra virksomhederne.

Det grundlæggende objekt ("stamenheden") i det erhvervsstatistiske registersystem er den juridiske enhed, dvs. den person eller det selskab, der som ejer eller bruger har ansvaret for aktiviteten. Af de godt 400.000 juridiske enheder, der er omfattet af erhvervsregistreringen, er ca. 85 pct. eller ca. 325.000 personligt ansvarlige virksomheder (firmaer), hvis ejere i registreringen er identificeret ved navn, adresse og personnummer.

De registrerede juridiske enheder kan udøve aktivitet inden for forskellige geografiske områder og inden for forskellige brancher, fx kan samme enkeltmandsvirksomhed udøve rådgivende ingeniørvirksomhed og drive landbrug.

Erhvervsregistreringen omfatter foruden oplysninger om de juridiske enheder også oplysninger om lokale fysiske enheder hørende under de ansvarlige juridiske enheder, fx filialer, afdelinger, fabrikker mv.

Det registersystem, der omtales her, må ikke forveksles med erhvervsregistret, som er et administrativt register med et mere begrænset indhold af oplysninger om de samme objekter. Erhvervsregistret administreres i medfør af økonomiministeriets bekendtgørelse nr. 580 af 25. nov. 1975 af Danmarks Statistik. Dets oplysninger er i et vist omfang offentligt tilgængelige i henhold til erhvervsregisterloven.

Det erhvervsstatistiske registersystem er opbygget med henblik på fremstilling af forskellige typer af erhvervsstatistik, og dets oplysninger er ikke tilgængelige for andre end Danmarks Statistik. For at sikre, at erhvervsregistrets oplysninger kan blive så korrekte og fuldstændige som muligt, fastsætter loven om erhvervsregistret og bekendtgørelsen om dette register, at der skal ske et samarbejde mellem erhvervsregistret og det erhvervsstatistiske registersystem om bestem-

te i loven opregnede arter af grundoplysninger.

Ved indsamlingen af oplysninger fra virksomhederne gør Danmarks Statistik opmærksom på, at de oplysninger, der er opregnet i erhvervsregisterloven, kan videregives til erhvervsregistret, og at de i dette register er omfattet af erhvervsregisterlovens videregivelsesregler. De oplysninger, der i øvrigt optages i det erhvervsstatistiske registersystem, er omfattet af de i Danmarks Statistik gældende diskretionsregler, der vedrører både personoplysninger og individualiserede oplysninger om virksomhederne.

Det erhvervsstatistiske registersystem består dels af en basisdel, som principielt omfatter identifikationsoplysninger om alle de erhvervsenheder, der anvendes i erhvervsstatistikken, dels af registreringer i delregistre vedrørende grupper eller udsnit af virksomheder inden for samme eller flere brancher.

2. Basisdelen

2.1. Enhedsbestand

Det erhvervsstatistiske registersystems basisdel omfatter principielt alle de objekter, der indgår i erhvervsregistret, dvs. erhvervsvirksomheder og arbejdsgivere, som er berettiget eller forpligtet til registrering hos offentlige myndigheder eller institutioner samt filialer eller afdelinger under sådanne. I alt omfatter registret som nævnt ca. 400.000 juridiske enheder og herudover ca. 50.000 særskilt beliggende enheder (fabrikker, filialer, institutioner).

2.2. Datatyper

I basisdelen indgår om de enkelte enheder oplysninger om enhedens navn, adresse, telefonnr., kommune, ansvarlig indehavers navn og personnummer, erhvervsregisternummer, juridisk organisationsform, art, herunder branche, samlet beskæftigelse, enhedens numre i skattevæsenets CIR-register, toldvæsenets VIRK-register og aktieselskabsregisternummer samt aktie- eller indskudskapital i aktie- og anpartsselskaber.

2.3. Anvendelse

Basisdelen benyttes dels ved udtagning af de udsnit af virksomheder, som skal indgå i de forskellige erhvervsstatistiske tællinger, dels som grundlag for statistiske opgørelser i samspil med oplysninger fra forskellige administrative registre. Basisde-

len benyttes endvidere ved vedligeholdelse af de erhvervsstatistiske delregistre, som omfatter bestande af virksomheder inden for de erhvervsområder, fra hvilke Danmarks Statistik indsamler erhvervsmæssige oplysninger til statistisk brug.

2.4. Opbygning og ajourføring

Ajourføringen af basisdelens registreringer om de juridiske enheder sker især på grundlag af oplysninger i toldvæsenets VIRK-register og skattevæsenets CIR-register og forudsætter derfor, at registret kan samkøres med disse donorregistre. Grundlaget for samkøringen er en aftale om fællesregler for registrering af juridiske enheder, der er truffet mellem blandt andre toldvæsenet, skattevæsenet og Danmarks Statistik. Fællesreglerne er imidlertid endnu ikke gennemført fuldt ud i donorregistrene, og den fælles enhedsnummerering, der er aftalt, er ikke iværksat. Problemerne omkring udbygningen af samarbejdet om registreringen af de juridiske virksomhedsenheder er genstand for løbende forhandlinger mellem registermyndighederne.

For så vidt angår registersystemets lokale enheder har mulighederne for systematisk ajourføring tidligere været utilstrækkelige. Gennemførelsen af arbejdspladsprojektet, der sker i et samarbejde mellem Danmarks Statistik og skattemyndighederne, har imidlertid nu skabt grundlag for en betydningsfuld udbygning og forbedring af denne del af erhvervsregistreringen. Som led i dette projekt rettes der hvert år henvendelse til arbejdsgiverne for at få ajourført oplysningerne om de lokale enheder, arbejdsstederne. Hovedsynspunktet ved opdelingen på lokale enheder er, at hver enhed skal være en selvstændig produktions- eller forretningsenhed omfattende aktiviteten på et enkelt sted (fabrik, værksted, butik, landbrug) under samme ejer. Krav om mere detaljerede statistiske opdelinger på geografiske områder fører til, at arbejdssteder, som ikke kan siges at udgøre selvstændige bedriftsenheder, men som beliggenhedsmæssigt er adskilt fra bedriftsenhederne, udskilles som specielt registrerede hjælpeenheder.

Både de juridiske basisenheder og de lokale enheder er branchebestemt efter hovedbranche og eventuelle bibrancher. Branchekodningen sker på grundlag af den internationale kode ISIC, der er udbygget under hensyn til danske forhold. I det mest detaljerede niveau anvendes 500-600 brancher.

I det erhvervsstatistiske registersystems basisdel indgår

bl.a. oplysninger om virksomhedernes navn og adresser fra toldvæsenets VIRK-register og skattevæsenets CIR-register.

Herudover omfatter registersystemets basisdel:

- MDS (momsdatasupplement indeholdende supplerende oplysninger fx om brancheforhold for VIRK-enhederne).
- ADS (arbejdsgiverdatasupplement indeholdende supplerende oplysninger fx om brancheforhold for CIR-enhederne).
- ASU (arbejdsstedssupplement indeholdende oplysninger om lokale registerenheder).

3. Det erhvervsstatistiske registersystems delregistre

I delregistrene registreres som nævnt oplysninger om de større eller mindre grupper af virksomheder, der indgår i de enkelte erhvervsstatistiske undersøgelser. Neden for anføres en fortegnelse over delregistre inden for systemet.

De oplysninger, der indsamles til brug for de forskellige statistikker, registreres på filer, som hver for sig indeholder et mere eller mindre begrænset antal typer af oplysninger. Ofte indgår oplysninger fra flere filer i en statistisk sammenhæng. Det kan i nogle tilfælde bero på et skøn, om en dataindsamling, som er optaget på flere filer, opfattes som et eller flere delregistre. Eksempelvis registreres oplysninger om eksport og import hver for sig. Da oplysningerne imidlertid i udenrigshandelsstatistikken indgår i en sammenhæng, betragtes registreringerne vedrørende im- og eksport i fortegnelsen som ét register.

I fortegnelsen er først anført delregistre, som omfatter virksomheder fra flere forskellige erhvervsområder. Senere i fortegnelsen opregnes delregistre til brug for opgørelser vedrørende eksempelvis industri, handel, bygge- og anlægsvirksomhed mv.

For hvert delregister anføres oplysninger om dataindholdet, størrelsen af enhedsbestanden og ajourføringshyppigheden.

Delregistre.Vedrørende virksomheder inden for flere erhvervsområder.Oplysninger, antal enheder, ajourføring.

- | | |
|---|---|
| 1. Løn mv. | Løn mv. meddelt på oplysnings-sedlerne til skattemyndighe-derne fra samtlige arbejds-gi-vere om samtlige ansatte. Ca. 200.000 enheder (årligt). |
| 2. ATP-bidrag | Oplysninger om indbetalte ATP-bidrag. Ca. 150.000 enheder (kvartalsvis) |
| 3. Momsafregninger | Udgående afgift, indgående af-gift, afgiftstilsvar og eks-portsalg for samtlige momsplig-tige. CA. 370.000 enheder (må-nedlig). |
| 4. Eksport og import | For hvert vareparti angivelse af værdi, vareart, mængde samt for importen, hvorfra varerne kommer og for eksporten, hvor-til varerne sendes. Ca. 40.000 importører (løbende). Ca. 25.000 eksportører (løbende). |
| 5. Driftsregnskabs- og statusop-lysninger for aktie- og an-partsselskaber | Hovedposter. Ca. 7.000 enheder (årligt). |
| 6. Mellemværender med udlandet | Gæld til og tilgodehavender i udlandet med angivelse af beløb, art samt om mellemvæ-renderne er kortfristede eller langfristede. Ca. 5.000 enhe-der (årligt). Udsendelsesregi-ster for de 300 største firma-er (kvartalsvis). |
| 7. Lønombkostninger | Antal udførte arbejdstimer og lønombkostninger. Ca. 5.000 en-heder (ca. hvert 3. år). |
| 8. Lønstrukturforhold | Oplysninger om beskæftigedes alder, køn, kvalifikation, an-ciennitet, arbejdssted, udfør-te arbejdstimer og aflønning. Ca. 5.000 enheder (ca. hvert 6. år). |
| 9. Funktionærlønninger | Oplysninger om beskæftigede funktionærs alder, køn, ud-dannelse, stillingsbetegnelse, anciennitet, udførte arbejds-timer og aflønning (ca. 10.000 enheder når statistikken er fuldt udbygget, årligt). |

10. Udenlandsk ejede firmaer

Firmaer her i landet, hvor udlandet har bestemmende indflydelse. Angivelse af det udenlandske moderselskabs navn og nationalitet samt andelen af aktie- eller anpartskapitalen i det danske firma. Ca. 1.600 enheder (ad hoc).

Vedrørende industri.

11. Beskæftigelse og løn

Antal beskæftigede indehavere, funktionærer, arbejdere, mænd og kvinder, hel- og deltid-beskæftigede, ungarbejdere og lærlinge. Antal udførte arbejdstimer. Udbetalt løn til arbejdere og funktionærer og hjemmearbejdere. Ca. 7.800 enheder (årligt). Ca. 1.500 enheder (månedlig).

12. Omsætning og ordreindgang

Omsætning, ordreindgang og ordrebeholdning for hjemme- og eksportmarked. Ca. 1.600 enheder (månedlig).

13. Varesalg-enkeltvarer

Samlet omsætning. Salg af egne varer. Specifikation på enkeltvarer på grundlag af NIMEXE-nomenklaturen med angivelse af mængde og værdi. Ca. 6.500 enheder (kvartalsvis).

14. Råvarekøb

Diverse rå- og halvfabrikata samt emballage. Ca. 3.000 enheder (ca. hvert 5. år).

15. Energiforhold

Fast og flydende brændsel, gas, elektricitet og fjernvarme. Lagerkapacitet, alternative energimuligheder og overskydende energi (arter). Rumopvarmning med procesenergi. Ca. 4.200 enheder (hvert 2. eller 3. år).

16. Investeringer

Fast ejendom, driftsmidler, forureningsbegrænsende foranstaltninger. Ca. 4.200 enheder (årlig).

17. Konjunkturforventninger

Forventning om mindre, uændret eller større produktion, beskæftigelse, samlet ordreindgang og investering. Bedømmelse af ordrebeholdning, eksportordrebeholdning, råvare og færdigvarelagre samt produktionskapacitet. Mangel på arbejdskraft, efterspørgsels- og kapa-

- citetsforhold som årsag til produktionsbegrænsninger. Ca. 1.000 enheder (kvartalsvis).
18. Forureningsbekæmpende foranstaltninger
Trufne tekniske foranstaltninger imod forurening, andre foranstaltninger til bortskaffelse af affald og spildprodukter. Ca. 2.000 enheder (ad hoc).
19. Driftsregnskabs- og statusoplysninger
Hovedposter fra driftsregnskaber og status. Ca. 3.500 enheder (årligt). Ca. 3.000 enheder (hvert 5. år).
- Vedrørende engroshandel.
20. Priser
Engrospriser for 1.000-1.500 enkeltvarer. Ca. 1.000 enheder (månedlig).
21. Driftsregnskabsoplysninger
Hovedposter fra driftsregnskaber. Ca. 5.500 enheder (årligt).
- Vedrørende detailhandel.
22. Omsætning
Månedlig omsætning i detailhandelsfirmaer. Ca. 1.000 enheder (månedlig). Ca. 6.000 enheder (hver anden måned).
23. Priser
Detailpriser for ca. 1.000 varer og tjenesteydelser. Ca. 1.000 enheder (månedlig).
24. Driftsregnskabsoplysninger
Hovedposter fra driftsregnskaber. Ca. 5.500 enheder (årligt).
- Vedrørende bygge- og anlægsvirksomhed.
25. Beskæftigelse
Antal beskæftigede mestre, medhjælpende hustruer, funktionærer, faglærte, ikke faglærte og lærlige. Ca. 30.000 enheder (hvert 3. år). Ca. 4.600 enheder (kvartalsvis).
26. Bruttolønsummer
Udbetalt bruttoløn til arbejdere og funktionærer. Ca. 2.400 enheder (kvartalsvis).
27. Konjunkturforsventninger
Forsventning om mindre, uændret eller større beskæftigelse og ordrebeholdning i det kommende kvartal. Oplysninger om arbejdskraftssituationen, herunder om eventuel mangel på fag-

- lærte eller ikke faglærte arbejdere. Ca. 650 enheder (kvar-
talsvis).
28. Driftsregnskabsoplysninger Hovedposter fra driftsregnska-
ber for bygge- og anlægsvirk-
somheder. Ca. 1.000 enheder
(årligt). Ca. 800 enheder
(hvert 5. år).
- Vedrørende transport.
29. Transportydelser på lastbiler Antal kørte km., godsets vægt,
art, på- og aflæsningssteder,
benyttelse af påhængsvogn. Ca.
1.000 vogne (kvar-
talsvis, skiftende udsnit).
30. Eksportvognmænd (Udsendelsesregister). Ca. 250
enheder (kvar-
talsvis).
31. Danske skibes fart på udlandet Fragts art, farvandsområde,
ladningens art samt fragtind-
tjening. Ca. 200 enheder (år-
ligt).
32. Driftsregnskabsoplysninger vedrørende rederier (Udsendelsesregister). Ca. 500
enheder (årligt).
- Vedrørende turisme.
33. Overnatninger i hoteller og på campingpladser Danske og udenlandske gæsters
overnatninger. Ca. 1.800 enhe-
der (månedlig).
34. Rejsebureau (Udsendelsesregister). Ca. 75
enheder (kvar-
talsvis).
- Vedrørende øvrige erhvervsområder.
35. Smede- og maskinreparations-
værksteder Beskæftigelse og omsætning.
Ca. 2.500 enheder (ad hoc).
36. Maskinstationer (Udsendelsesregister). Ca.
1.300 enheder (ad hoc).
37. Kornhandlere (Udsendelsesregister). Ca. 640
enheder (tre gange hvert år).
38. Foderstofhandlere (Udsendelsesregister). Ca. 180
enheder (halvårligt).
39. Producenter af foderblandinger (Udsendelsesregister). Ca. 410
enheder (halvårligt).
40. Skove Skovenes areal, art, alder,
bonitet, ejerforhold. Ca.
8.000 enheder (ad hoc).

41. Landbrugskonsulenter

(Udsendelsesregister). Ca. 150
konsulenter (årligt).

Bilag 10: Litteraturhenvisninger

- (1) Andersen, Tavs Folmer: Børns opvækstvilkår i langsigtet perspektiv. Institut for Social Medicin, Københavns Universitet, 1981.
- (2) Betænkning afgivet af folkeregisterkommissionen af 1920, København 1922.
- (3) Betænkning om folkeregistrenes medvirkende ved indførelse af elektronisk databehandling i den offentlige forvaltning mv. Betænkning nr. 329. København 1963.
- (4) Danmarks Statistik: Befolkningens bevægelser 1980. Statistiske meddelelser 1982:1, København 1982.
- (5) Danmarks Statistik: Boligtællingen 1. januar 1980. Statistisk tabelværk 1981:III. København 1981.
- (6) Danmarks Statistik: Dødelighed og erhverv 1970-75. Statistiske Undersøgelser nr. 37. København 1979.
- (7) Danmarks Statistik: Registerfolketællingen RF76. Statistisk Tabelværk 1978:VIII, 1979:I, 1979:II, 1981:IV. København 1978-81.
- (8) Delbetænkning om offentlige registre. Betænkning nr. 767. København 1976.
- (9) National Center for Health Statistics: The Person-Number systems of Sweden, Norway, Denmark, and Israel. Data Evaluation and Methods Research, Series 2, Number 84. Hyattsville, Maryland 1980.
- (10) Nordbotten, Svein: Om arkivstatistiske systemer. Optaget som bilag 4 i Nordisk Statistisk Skriftserie nr. 14, Nordisk Statistisk Sekretariat. København 1967.
- (11) Nordisk Statistisk Sekretariat: Censusproblematikken. Kap. 2 i Nordisk Statistisk Skriftserie nr. 32. København 1977.
- (12) Nordisk Statistisk Sekretariat: Rapport om ett system för socio-demografisk statistik (SSDS). Tekniske rapporter nr. 22. København 1979.
- (13) Lov om offentlige myndigheders registre. Lov nr. 294 af 8. juni 1978.
- (14) Lov om private registre mv. Lov nr. 293 af 8. juni 1978.
- (15) Statistiska Centralbyrån: Produktionstekniska principer vid SCB. Stockholm 1979.

- (16) Sundgren, Bo: An Infological Approach to Data Bases. URVAL nr. 7, SCB, Stockholm 1973.
- (17) Sundgren, Bo: Infologisk udformning af statistiske databaser. Statistisk Tidsskrift 1975:4. Stockholm 1975.
- (18) UN: Methods of Collecting, Organizing and Retrieving Social Statistics to Achieve Integration (E/CN.3/516). New York, 1978.
- (19) US Department of Commerce: Report on Exact and Statistical Matching Techniques. Statistical Policy Working Paper 5. Washington D.C., 1980.
- (20) US Department of Commerce: Report on Statistical Uses of Administrative Records. Statistical Working Paper 6. Washington D.C. 1980.