

# Interne notater

STATISTISK SENTRALBYRÅ

92/7

juni 1992

## DATAGRUNNLAGET FOR TRYGDEMODELLEN MOSART-T.

Av Dennis Fredriksen.

### Sammendrag

Notatet dokumenterer datagrunnlaget for trygdemodellen MOSART-T, til bruk i analyser for og programmering av modellen. Andre viktige bruksområder er ved andre analyser/statistikkproduksjon og ved framtidige oppdateringer av datagrunnlaget. Notatet er derfor relativt detaljert. EDB-programmer er likevel holdt utenfor, og disse refereres til i et upublisert notat.

### Innhold

1. Innledning. ....	3
2. Utvalget. ....	5
3. Flyttehistorier. ....	11
4. Ekteskaphistorier. ....	12
5. Fødselshistorier. ....	13
6. Utdanningshistorier. ....	14
7. Trygde- og yrkeshistorier. ....	16
8. Tilrettelegging av data. ....	18
Vedlegg. ....	23
A. Filbeskrivelser, Mellomresultater. ....	24
B. Filbeskrivelser, MOSART. ....	27
C. Filbeskrivelser, sekvensielle filer. ....	29
D. Formater. ....	32



## 1. Innledning.

MOSART<sup>1</sup> er en empirisk mikrosimuleringsmodell som bygger på den gode tilgangen på registerstatistikk i Norge. Modellen starter med et utvalg av befolkningen, en såkalt utgangspopulasjon, og simulerer det videre livsløpet for hvert enkelt individ i dette utvalget. Førsteutgaven av MOSART simulerer begivenheter som inn- og utvandring, dødelighet, bevegelser i ekteskapelig status, fødsler, skolegang og arbeidstilbud. Simuleringen bygger på overgangssannsynligheter som angir sannsynligheten for at en eller flere av disse begivenhetene skal inntreffe i et gitt år for hvert individ. Modellen trekker så på grunnlag av dette hvor hvert individ går, og når dette gjøres for tilstrekkelig mange individer, kan man si noe om befolkningen. Nye årskull av 16-åringer legges til etterhvert, og resultatet av simuleringen blir en modellpopulasjon av individer med sine tildels historiske og tildels simulerte livsløp. Når modellpopulasjonen er representativ for befolkningen, vil man gitt overgangssannsynlighetene i modellen, kunne gi framskrivninger av befolkningen for den perioden man simulerer. En foreløpig dokumentasjon av førsteutgaven av MOSART finnes i Andreassen og Fredriksen (1991) og i statistisk ukehefte 91/1&2.

Mye av den input som inngår i MOSART er basert på tidligere utførte arbeider, og disse har ofte svakheter i forhold til krav MOSART stiller. Vi har derfor tilrettelagt modellen for videreutvikling og forbedringer. Dette arbeidet er en del av et prosjekt som skal videreutvikle MOSART til en modell som kan analysere alders- og uføretrygd fra Folketrygden. Motivasjonen for dette prosjektet er at tilleggspensjonen i Folketrygden er avhengig av individuelle yrkeshistorier, og at arbeidstilbudet i aldersgruppene 55-66 år er sterkt påvirket av omfanget av uføretrygd. Vi tror derfor MOSART er velegnet til å fange opp disse problemstillingene. På datasiden vil prosjektet bestå i å koble Rikstrygdeverkets registre for opptjente pensjonsrettigheter mot SSB's registre for demografi og utdanning. På estimeringssiden vil et arbeidstilbud som gir yrkeshistorier med arbeidsinntekt være av stor betydning. I tillegg vil også overganger til uførhet være av betydning. Programmeringen vil i stor grad bygge videre på førsteutgaven av MOSART. En mer utførlig presentasjon av prosjektet med å utvikle MOSART til en trygdemodell finnes i Fredriksen (1991).

En viktig del av å videreutvikle MOSART til en trygdemodell består i å framskaffe bedre mikrodata. I utgangspopulasjonen er det nødvendig å kjenne de pensjonsrettigheter som dette utvalget hittil har opptjent. Gitt beregningsreglene for alders- og uføretrygd, vil disse opptjente rettighetene i stor grad bestemme utbetalingene til alders- og uføretrygd i mange år framover. I tillegg er det ønskelig å framskaffe mikrodata som kan brukes til å estimere overgangssannsynligheter knyttet til arbeidsinntekt og overganger til uførhet. Spesielt gjelder dette individuelle inntektsdata som strekker seg over lengre tidsrom. Den omtalte datasiden av dette prosjektet sikrer databehovet for

---

<sup>1</sup> MOSART er et akronym for "MOdell for mikrosimulering av Skolegang og ARbeidsTilbud".

utgangspopulasjonen og estimeringsarbeidene. Dette notatet tar opp dette dataarbeidet og gir en dokumentasjon av resultatfilene. Siktemålet er at eventuelle brukere av datagrunnlaget skal kunne bruke dataene på en kritisk måte. Videre skal dokumentasjonen kunne tjene som en brukerveiledning ved oppdateringer og utvidelser av datagrunnlaget for MOSART. Omtalen blir derfor på enkelte punkter detaljert. EDB-programmer for filkoblingene blir derimot bare løselig omtalt her. Disse programmene er skrevet i SAS for SSB's Comparex-maskin, og er lite informative for andre enn aktive SAS-brukere. Disse finnes imidlertid listet ut i ett eget upublisert notat, Fredriksen (1992), og blir referert i teksten.

Kapittel (2) gjennomgår trekningen av utvalget. Kapitlene (3-7) gjennomgår hvordan data er hentet inn for de ulike kjennetegnene. Kapittel (8) gjennomgår hvordan utvalget og dataene har blitt tilrettelagt for analyseformål og annen bruk. Kapitlene (2-8) innledes med en mer generell omtale, før eventuelle detaljer knyttet til EDB-arbeidet tas opp. Vedlegg (A-C) gir en beskrivelse av de viktigste resultatfilene. Vedlegg (D) gir en beskrivelse av de variable som inngår i datamaterialet (verdier, forkortelser).

Ved koblinger av registerdata er det viktig å legge stor vekt på personvern og ta hensyn til dette ved planleggingen og gjennomføringen av prosjektet. Vi vil regne med at fødselsnummeret vil bli erstattet av et såkalt kryptert fødselsnummer i alle registre. Dette vil ha den fordel at sammenhengen mellom person og fødselsnummer blir fullstendig en-entydig. Derimot vil kobling av registre med ulike krypteringsnøkler blir vanskeligere, spesielt på grunn av omkrypteringer og omfattende sorteringer. Søknader til registereiere og Datatilsynet bør derfor sendes i god tid, gjerne opp mot ett år før EDB-arbeidet er planlagt å begynne. I denne sammenheng vil det si Sosialdepartementet som eier av Rikstrygdeverket's pensjonpoengregister som kanaliseres gjennom seksjon for helse, sosial og trygd hver høst. Søknad til Datatilsynet sendes direkte av prosjektet/avdelingen, men bør rådføres med sikkerhetsutvalget eller lignende.

## 2. Utvalget.

Tilretteleggingen av data for førsteutgaven av MOSART krevde uventet mye ressurser og tid. Noe av årsaken til dette var mengden data som inngikk. Dette dataprojektet omfatter enda flere data fra flere kilder. Vi tok derfor tidlig den beslutningen at filkoblingene kun skulle gjennomføres for et utvalg av befolkningen. Innledningen til dette kapitlet gir en generell beskrivelse av hvordan dette utvalget er trukket og noen av de statistiske egenskaper dette utvalget har. Figur (1) på neste side gir en oversikt over sammenhengen mellom de EDB-programmene som er brukt for å trekke utvalget.

Utvalget er uveid forhåndsstratifisert etter kjønn, alder, ekteskapeleg status, reg.status<sup>2</sup> og årstall for reg.status. Utvalget er uveid fordi hver person i simuleringsmodellen skal kunne tillegges samme vekt ved oppblåsning av tallene. Utvalget er forhåndsstratifisert for å redusere variansen og korrigere for mulige skjevheter. Spesielt vil aldersstrukturen via dødelighet påvirke befolkningsutviklingen, og variansen knyttet til dette er eliminert ved å stratifisere etter alder. En del personer er utelukket fra trekningen fordi opplysninger om ekteskapeleg status er mangelfulle, og disse personene er skjevt fordelt etter ekteskapeleg status og alder. Dette blir (delvis) rettet opp i stratifiseringen.

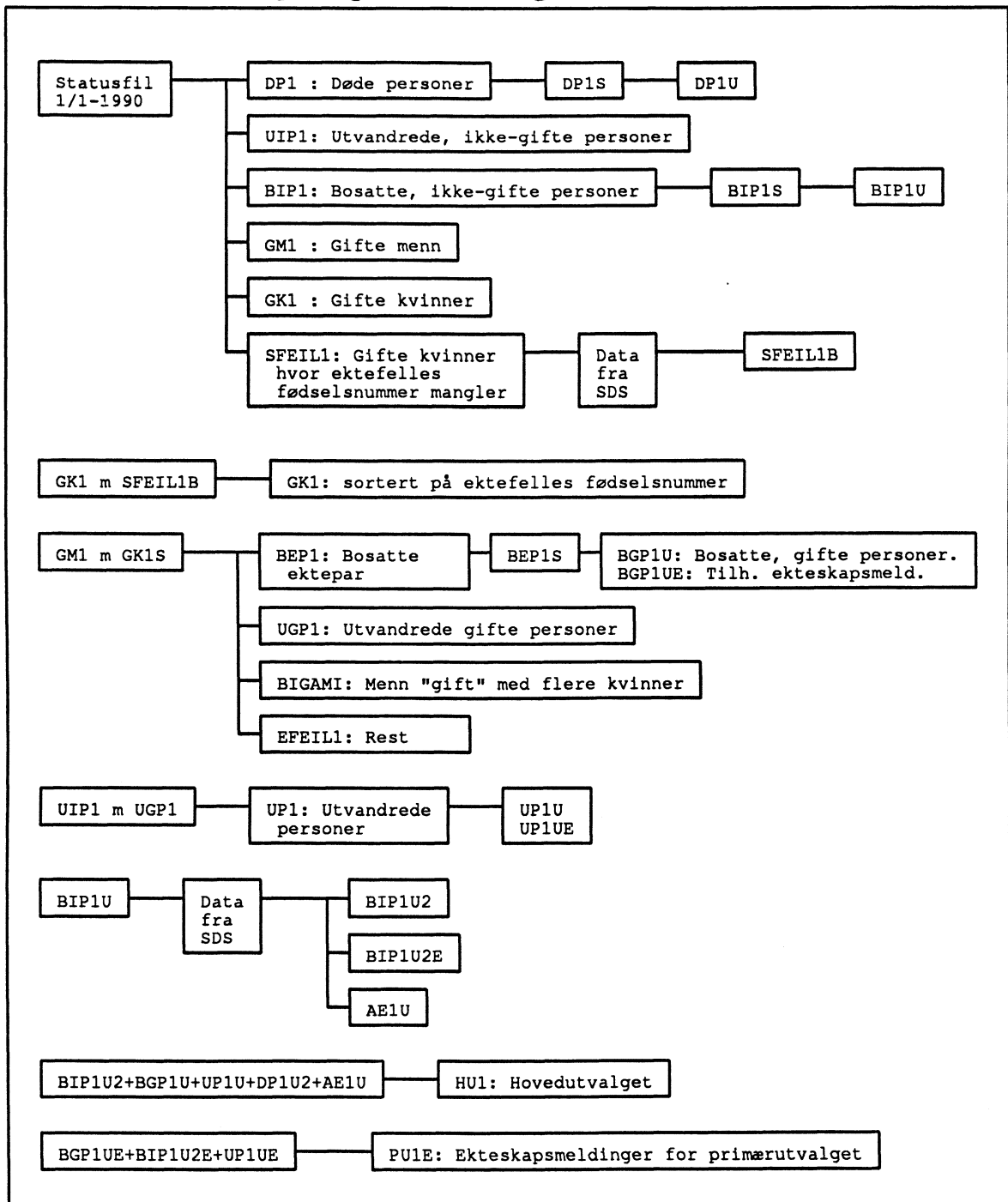
Siden utvalget er stratifisert etter ekteskapeleg status, er opplysninger om disse forholdene innhentet ved trekningen av utvalget. Utvalget omfatter bare personer som er 16 år og eldre ved utgangen av 1989. Første del av utvalget omfatter 10 prosent av befolkningen pr 1/1-1990, hvor ektefeller er trukket sammen. Første del av utvalget danner grunnlaget for utgangspopulasjoner i simuleringsmodellen. Andre del av utvalget utgjør 10 prosent av personer som har utvandret eller dødd i perioden 1967-1989. Disse er nødvendige hvis man skal analysere overganger i denne perioden som er korrelerte med utvandring eller dødelighet. Et eksempel er hvis uføretrygdede har høyere dødelighet, da vil de som har blitt uføretrygdet før 1990 være underrepresentert i befolkningen pr 1/1-1990. Tilsammen utgjør første og andre del av utvalget det såkalte primærutvalget, og vil bli brukt ved estimering av overgangssannsynligheter og for å ta ut historiske tidsserier for de størrelser modellen framskriver.

En viktig egenskap ved utgangspopulasjonen er at alle individer skal tillegges samme vekt ved oppblåsning av tallene. Dette kan sikres ved å trekke et tilfeldig utvalg individer, eventuelt bruke en uveid forhåndsstratifisering. Det er også mulig å dele befolkningen i klynger, slik at hvert individ tilhører en og bare en klynge, og så trekke et tilfeldig utvalg klynger. Dette har vi benyttet i MOSART ved å bruke ektepar (på et gitt tidspunkt) som klynge. Alternativt kunne vi benyttet husholdning på tilsvarende måte. Mer kompliserte relasjoner enn dette er neppe hensiktsmessig å bruke, da man enten vil få at noen

---

<sup>2</sup> Reg(istrerings)status vil si om man er bosatt i Norge, er død eller har utvandret. Bosatt i Norge forkortes ofte til "bosatt".

**Figur 1. Trekking av primærutvalget.**



personer vil tilhøre flere klynger eller at alle indirekte er i "slekt" med hverandre<sup>3</sup>.

<sup>3</sup> Et eksempel kan være foreldre-barn. Tar man med alle barn, vil antallet klynger en person tilhører avhenge av antall foreldre/besteforeldre/osv som er i live, og dette varierer fra individ til individ. Tilsvarende blir det med foreldre, her vil antallet levende etterkommere variere enda sterkere. Tar man toveis-relasjonen foreldre-barn vil man nøste opp hele befolkningen fordi man må ta med alle søsken, alle

Derimot kan det være av interesse å få med informasjon om andre relaterte personer enn ektefellen. I denne utgaven av trygdemodellen har vi begrenset oss til avdøde ektefeller<sup>4</sup> av enker og enkemenn som er bosatt pr 1/1-1990 (og som kom med i første del av utvalget). Disse avdøde ektefellene kan ha opptjent pensjonrettigheter som kan påvirke tilleggspensjonen for gjenlevende ektefelle. Disse supplerende personene utgjør sammen med primærutvalget det såkalte hovedutvalget. Det er for hovedutvalget vi henter inn registeropplysninger for opptjente pensjonsrettigheter i Folketrygden, visse demografiske begivenheter og utdanning.

I tillegg er hovedutvalget delt opp i ti (tilfeldige) underutvalg på en prosent. Disse vil være egnet som utgangspopulasjoner i simuleringsmodellen. I tillegg kan de brukes som mindre utvalg for å redusere forbruket av regnekraft, spesielt ved utvikling av mulige økonomiske opplegg. Ved helt passiv søking etter statistiske sammenhenger, kan ubenyttede deler av utvalget også brukes til å kontrollere om de sammenhenger man har funnet er mer generelle.

## 2.a. Innlesning av statusfil.

Den såkalte statusfilen som SSB besitter omfatter alle personer som har vært tildelt et personnummer i Norge, og omfatter en rekke opplysninger fra det Sentrale personregister. Programmet (A.1. Statfil) leser inn hele statusfilen med de kjennetegn som er aktuelle i denne koblingen:

- Fødselsnummer,
- fødselsdato,
- kjønn,
- reg.status,
- dato for reg.status,
- ekteskapeleg status,
- dato for ekteskapeleg status og
- eventuell ektefelles fødselsnummer.

Dataene fordeles på syv disjunkte datasett:

- DP1 ; Døde personer pr 1/1-1990.
- BIP1 ; Ikke-gifte personer som var bosatt i Norge pr 1/1-1990.
- UIP1 ; Ikke-gifte personer som var utvandret fra Norge pr 1/1-1990.
- GM1 ; Gifte menn pr 1/1-1990.
- GK1 ; Gifte kvinner pr 1/1-1990, i tillegg opprettes undermengden SFEIL1 med gifte kvinner som mangler ektefelles fødselsnummer.
- SFEIL2; Ikke-døde per 1/1-1990 med ukjent ekteskapeleg status (få personer).
- <> ; Personer med ukurant reg.status (få personer).

---

personer søsknene har barn med, alle søsken av disse, osv. Subsidært kunne man brukt relasjonen mor-barn, men denne kan ikke kombineres med ekteskap/husholdning.

<sup>4</sup> Disse kunne i prinsippet ha vært trukket ut i primærutvalget via sine gjenlevende ektefeller. Dette krever at det direkte utvalget av døde ikke omfatter avdøde ektefeller, og dette har vi ikke muligheten til å kontrollere i denne versjonen da ekteskaphistorier er for vanskelig tilgjengelig.

## **2.b. Separerte kvinner.**

På statusfilen mangler opplysninger om tidligere ektefeller for før gifte personer, inklusive separerte. MOSART behandler separerte som "gifte" og disse opplysningene må derfor innhentes ved en spesialkjøring i Statens datasentral<sup>5</sup>. En søknad må rettes til Sentralkontoret for folkeregistrering, og en behandlingstid på 2-6 uker må påberegnes. Programmene (A.2a. SPKVDISK/A.2b. SPKVTAPE) leser separerte kvinner fra datasettet SFEIL1 ut på en sekvensiell fil som sendes Statens Datasentral. I retur fås blant annet fødselsnummeret for ektefellen til separerte kvinner, og programmet (A.2c. SPKVDATA) leser dette inn som SAS-datasettet SFEIL1B.

## **2.c. Kobling av ektefeller.**

Programmet (GK1SORT) leser opplysningene fra SFEIL1B inn på datasettet GK1, og sorterer disse dataene på ektefellens (mannens) fødselsnummer, og legger resultatet ut på datasettet GK1S. Programmet (A.4. EP1DATA) slår sammen datasettene GM1 og GK1S på grunnlag av mannens fødselsnummer og legger resultatet ut på fire disjunkte datasett:

- BEP1; Ektepar hvor begge ektefeller er bosatt i Norge og er 16 år og eldre.
- UGP1; Gifte personer som har utvandret.
- EFEIL1; Bosatte gifte personer hvor ektefellen har utvandret, ikke finnes eller er ukjent. Omfatter anslagsvis 20 000 personer.
- BIGAMI; Tilfeller hvor flere kvinner er gift med samme mann, og skyldes at ekteskapelig status/ektefelle ikke oppdateres for utvandrere.

NB! På grunn av feil i programmeringen har det ikke blitt sjekket at begge ektefeller er 16 år og eldre. I denne utgaven har ikke dette vært noe problem da ingen gifte er så unge. I tillegg er fødselsdato for gifte/separerte kvinner som har utvandret, blitt uteglemt (dette er rettet opp i programmet senere). Dette er også rettet opp på filen i en senere databearbeiding, men det er glemt å sjekke for århundre slik at kvinner født i forrige århundre som er gift eller separert, og utvandret, har falt ut.

## **2.d. Tilrettelegging av utvandrere.**

Gifte utvandrere ble i 2.c. brukt til å kontrollere hvor mange bosatte gifte personer som har ektefellen bosatt i utlandet. Med programmet (A.5. UP1DATA) blir ikke-gifte og gifte utvandrere slått sammen til datasettet UP1.

## **2.e. Trekkeprosedyrer.**

Disse er bygd opp i to trinn hvor første steg sorterer datasettet på de variablene trekningen skal forhåndsstratifiseres etter. Steg to leser så inn utvalgsplanen som viser hvor mange personer som finnes i hver gruppe i totalbefolkningen og hvor mange som skal trekkes i hver slik gruppe. Deretter

---

<sup>5</sup> Statens datasentral tar betalt for denne typen oppdrag, og prisen bør sjekkes før oppdraget startes.



trekkes det hypergeomtrisk innen hver slik gruppe<sup>6</sup>. Utvalgsplanen fordrer tabeller som både beskriver befolkningen og den totalpopulasjonen det trekkes fra. Hver trekkeprosedyre resulterer i to datasett, hvor det ene omfatter personene som er trukket ut og det andre omfatter ekteskapsmeldinger for disse personene basert på ekteskapsstatus pr 1/1-1990.

Programmene (B.1a. DP1SORT/B.1b. DP1TRK) sorterer datasettet med døde personer (DP1) etter kjønn og dødsår og trekker et 10-prosent utvalg i forhold til dødelighetsstatistikk for årene 1967-1989. Utvalget legges ut på datasettet DP1U, mens ekteskapsmeldingen utelates da disse opplysningene trolig er av meget dårlig kvalitet.

Programmene (B.2a. BIP1SORT/B.2b. BIP1TRK) sorterer datasettet med bosatte ikke-gifte personer (BIP1) etter kjønn, ekteskapsstatus og fødeår, og trekker et 10-prosent utvalg i forhold til befolkningsstatistikk pr 1/1-1990. Utvalget legges ut på datasettet BIP1U, mens arbeidet med ekteskapsmeldingene utsettes til punkt (2.F) for å få med avdøde ektefeller av enker og enkemenn.

Programmene (B.3a. BEP1SORT/B.3b. BEP1TRK) sorterer datasettet med bosatte ektepar (BEP1) etter kvinnens ekteskapsstatus og fødeår, og trekker et 10-prosent utvalg i forhold til befolkningsstatistikk pr 1/1-1990. Utvalget legges ut på datasettet BGP1U, og ekteskapsmeldingene på datasettet BGP1UE, og hver av ektefellene representeres med en observasjon (bosatte gifte personer).

Programmene (B.4. UP1TRK) trekker et ikke-stratifisert 10-prosent utvalg av datasettet med utvandrere (UP1), da vi ikke kjenner gjeninnvandring blant de som utvandret i hvert enkelt år. Utvalget legges ut på datasettet UP1U og ekteskapsmeldingene ut på datasettet UP1UE. Noen personer som har utvandret før 1967 har også kommet med i utvalget (ved en glipp), men dette er så få at det ikke er rettet opp i denne versjonen.

## **2.f. Avdøde ektefeller.**

På statusfilen mangler opplysninger om tidligere ektefeller for før gifte personer. Tilleggspensjonen i Folketrygden kan avhenge av pensjonrettigheter en avdød ektefelle har opptjent. Opplysninger om hvem disse avdøde ektefellene er må derfor innhentes ved en spesialkjøring i Statens datasentral<sup>7</sup>. En søknad må rettes til Sentralkontoret for folkeregistrering, og en behandlingstid på 2-6 uker må påberegnes. Programmene (C.1a. EEDISK/C.1b. EETAPE) leser utvalget av enker og enkemenn fra datasettet BIP1U ut på en sekvensiell fil som sendes Statens Datasentral. I retur fås blant annet fødselsnummeret for siste ektefelle til enker og enkemenn i utvalget, og programmet (C.1c. EEDATA)

---

<sup>6</sup> Det vil si trekning uten tilbakelegging og slik at antall personer i utvalget blir eksakt lik det i utvalgsplanen. Dette sikres ved at trekkeprogrammet tar for seg en og en observasjon, og justerer trekkesannsynligheten for de gjenværende personene, avhengig av hvor mange disse er og hvor mange som gjenstår å trekke.

<sup>7</sup> Statens datasentral tar betalt for denne typen oppdrag, og prisen bør sjekkes før oppdraget startes.

leser dette inn som SAS-data som kobles med BIP1U. Utvalget legges så ut på datasettet BIP1U2 og ekteskapsmeldingene legges ut på datasettet BIP1U2E. I tillegg legges avdøde ektefeller ut på datasettet AE1U. En god del ektefeller mangler her, men dette dreier seg i stor grad om kvinner som ble enker for mange år siden. Man kan derfor anta at disse i liten grad har fått tilleggs-pensjon på grunnlag av mannens opptjente pensjonrettigheter.

NB! Ved beregning av fødselsdato har vi ikke sjekket for århundre (dette er rettet opp i programmet senere, men ikke kjørt på nytt). Det betyr at avdøde ektefeller født i forrige århundre har falt ut. Dette betyr trolig lite da de i meget begrenset grad har hatt muligheten til å opptjene pensjonsrettigheter.

## **2.e. Hovedutvalget.**

Programmet (C.2. HU1DATA) slår sammen de ulike delutvalgene:

- BIP1U2 ; Bosatte ikke-gifte personer.
- BGP1U ; Bosatte gifte personer.
- UP1U ; Personer som har utvandret.
- DP1U ; Døde personer.
- AE1U ; Avdøde ektefeller av enker og enkemenn.

til et felles datasett (HU1) med en observasjon for hver forekomst<sup>8</sup>. I tillegg påføres en indikator for om observasjonen kommer fra primærutvalget eller ikke. Denne filen markerer en klar milepæl i koblingsarbeidet og brukes i mange ulike anledninger for å hente inn data. Filen er derfor dokumentert i vedlegg (A).

## **2.e. Ekteskapsmeldinger.**

Programmet (C.3. PU1EDATA) slår sammen de ulike ekteskapsmeldingene:

- BIP1U2E ; Bosatte ikke-gifte personer.
- BGP1UE ; Bosatte gifte personer.
- UP1UE ; Personer som har utvandret.

til et felles datasett (PU1E) med en observasjon for hver melding. Denne filen representerer et av sluttproduktene og er derfor dokumentert i vedlegg (A).

---

<sup>8</sup> Dette betyr at noen personer kan forekomme to ganger, både som primærobservasjon og som avdød ektefelle.

### 3. Flyttehistorier.

Ved uttak av tidsserier og ved estimeringer er det viktig å vite om personen er bosatt i Norge<sup>9</sup>. Dette er ikke lett tilgjengelig informasjon, og dataene baserer seg derfor på litt sviktende grunnlag. Den ene kilden er reg.status pr 1/1-1990 og når denne sist ble endret, og er oppsummert i kjennetegnet "vital". Vi har antatt at personer som har dødd eller utvandret, i sin helhet har vært bosatt i Norge før dette, og at innvandrere ikke har vært bosatt i Norge før siste innflytting. Dette vil gi visse skjevheter, dess mer jo lengre tilbake vi kommer. Foreløpig virker totaltallene rimelige med unntak av for mange personer rundt 20 år (om lag 40 000 i 1973, og siden gradvis færre). Årsaken til dette er ikke oppklart.

Fra BHU-filene fra årene 1985-1989 har vi også hentet (formell) bostedskommune. Denne bør gi noe bedre bakgrunn, men er foreløpig ikke kontrollert. I tillegg kan bostedskommune brukes til analyse av regionale forskjeller. Ved analyse av flyttehistorier bør man imidlertid være mer kritisk da variabelen i liten grad er sjekket for feil. Et annet problem er personer som har fått endret fødselsnummer, da disse vil mangle de deler av historien som kom før siste endring. Dette er trolig et lite problem. Det er i denne versjonen ingen separate programmer for flyttehistorier.

NB! NB!

Ved bruk av individhistoriefilen (se kapittel 8), er det viktig at man tester både mot dublett (er personen med i primærutvalget) og mot vital (er personen bosatt i Norge).

---

<sup>9</sup> "Personer under risiko".

#### 4. Ekteskapshistorier.

Ekteskapeleg status og ektefelle inngår som en viktig del av modellen MOSART og det har vært viktig å få dette med for utgangspopulasjonen. I en trygdemodell er dette spesielt viktig da en etterlatt ektefelle kan ha krav på tilleggspensjon basert på avdødes pensjonspoeng. I tillegg ønsker vi å bruke ekteskapeleg status som kovariat i modellen som skal beregne overgangssannsynlighetene og muligens også se på ektefellers simultane tilpasning i arbeidsmarkedet. Sistnevnte vil kreve fulle ekteskapshistorier, noe som i prinsippet finnes tilgjengelig. Imidlertid er disse ekteskapsmeldingene spredd over et stort antall filer med ulik oppbygging og med et ikke ubetydelig innslag av inkonsistenser.

Vi har derfor nøydt oss med ekteskapeleg status pr 1/1-1990 og når denne status ble opprettet, slik disse opplysningene framkommer ved trekningen i utvalget. Det er da fullt mulig og bruke ekteskapeleg status som kovariat, eventuelt se på ektepar, for året 1989. Før dette blir dataene av stadig dårligere kvalitet, og man bør være meget varsom da de som har endringer i ekteskapeleg status kan være et meget skjevt utvalg av befolkningen. Spesielt kan ikke datagrunnlaget brukes som grunnlag for analyse av bevegelser i ekteskapeleg status.

##### 4.a. Ekteskapshistorier.

Programmet (D.3. EKTEHIST) tar ekteskapsmeldingene (PU1E) som kommer fra trekningen av primærutvalget og fletter ("merge") disse med stammen (se kapittel 8) etter fødselsnummer og årstall. Meldingene ligger sortert etter fødselsnummer og dato for begivenheten, og bare siste melding for hvert år er tatt med (unødvendig i denne versjonen). Hver gang en ekteskapsmelding påstøtes oppdateres ekteskapeleg status og ektefelle, og en endringsindikator for ekteskapeleg status legges til. Ellers settes ekteskapeleg status og ektefelle til det samme som for fjoråret. Hvis en ny person hentes opp uten ekteskapeleg status, settes ekteskapeleg status til ugift hvis personen er det pr 1/1-1990 (utgangen av startåret), ellers til uoppgitt. Stammen vil også inneholde ekteskapeleg status pr 1/1-1990 for hvert eneste år på individhistoriefilen.

NB! NB!

På de ekteskapsmeldinger hvor årstall for endring mangler, har vi forutsatt at endringen fant sted før 1967 (egentlig 1964).

## 5. Fødselshistorier.

Fødsler inngår også som en viktig del av modellen MOSART, både som kovariat for kvinners atferd, spesielt fertilitet, og som en mulighet for å bestemme befolkningstilveksten i modellen. Datagrunnlaget må da spesielt gi befolkningen 0-15 år pr 1/1-1990 (startåret). I tillegg ønsker vi å bruke barn som kovariat for kvinners overgangssannsynligheter. Basis er Personregisteret som har registrert alle fødsler etter 1952. Før dette vil fødsler i stadig mindre grad være registrert. Det betyr at vi til en viss grad får med fødselshistorien for kvinner født etter 1920, og i sin helhet for kvinner født etter 1935. Opplysningene er tilrettelagt gjennom Kvinnefilen, hvor alle fødsler<sup>10</sup> er samlet. Dette gjør datagrunnlaget egnet til å bruke barn som kovariat for hele perioden, samt til analyse av fertilitet.

### 5.a. Fødselshistorier.

Programmet (D.1. UTVALG) henter utvalget av kvinner fra utvalget (HU1) og programmet (D.2. KVINFIL) bruker dette til å hente utvalget av fødselsmeldinger fra Kvinnefilen. Disse ligger sortert på kvinnens fødselsnummer og barnets fødselsdato, og der flere fødsler (barn) forekommer i ett år, er disse samlet til en melding med "flerfødsel". Programmet (D.3. UNGDOM) henter reg.status for de av barna som er ikke-gift pr 1/1-1990 (startåret), og det vil i praksis inkludere alle barn under 16 år. Programmet (D.4. FODSHIST) fletter ("merge") disse meldingen med stammen (se kapittel 8) etter fødselsnummer og år for begivenhet. Variable som legges inn er antall barn og yngste barn's alder ved utgangen av året.

---

<sup>10</sup> Inkluderer adopsjoner.

## 6. Utdanningshistorier.

Utdanning er også en sentral del av modellen MOSART, og det er derfor viktig å få utdanningsstatus for utgangspopulasjonen. Videre ønsker vi å bruke utdanning som kovariat for annen atferd i modellen, spesielt arbeidstilbud. I tillegg mangler vi data som kan belyse utdanningsatferd over flere år sett i sammenheng med andre kjennetegn. Det har derfor vært et poeng å få med utdanningshistorier så langt tilbake som mulig. Problemene med dette er endringer i utdanningsstandarden og at BHU-filene ikke finnes for alle år før 1985. Av den grunn ville det være for ressurskrevende for dette prosjektet å få med alle utdanningskjennetegn før 1985. Vi har derfor bare tatt med utdanningsnivå etter ny standard for årene 1985-1989. Utdanningsaktiviteter har vi fått med for årene 1974-1989, men med gammel standard fram til og med 1984 og ny standard fra og med 1985.

Kvaliteten er ennå ikke kontrollert, men datamaterialet bør være egnet til å belyse effekten av utdanning på (andre) atferdsmønstre for årene 1985-1989. Før dette bør man være mer forsiktig, kanskje med unntak av overgang til uførhet. Dette gjelder i stor grad eldre mennesker, og disse har i liten grad endret utdanningsnivå etter 1967.

### 6.a. Innlesning av BHU-filer.

Programmet (E.1. BHUFIL) leser inn en og en årgang BHU-fil og plukker ut de individer som finnes i hovedutvalget. Dataene samles på disk og arbeidet bør av den grunn ikke dra ut i tid. Høyeste fullførte utdanning tas med som et mål for utdanningsnivå, med utdanningens art (6-sifret), klassetrinn (2-sifret) og fullføringstidspunkt (år-måned). Igangværende utdanning tas med som mål for utdanningsaktiviteter, med utdanningens art (6-sifret) og klassetrinn (2-sifret). I tillegg tas med (formell) bostedskommune.

### 6.b. Utdanningsdata I.

Programmet (E.2. UTDDATA) fletter stammen (se kapittel 8) med utdraget av de fem årgangene BHU-filer, etter fødselsnummer og årstall. En dummyvariabel "student" legges til for årene 1985-1989 hvor '1' angir at personen ikke under utdanning og '2' angir at personen er under utdanning. Programmet brukte urimelig mye CPU-tid, og dette kunne vært betydelig redusert ved å flette sammen BHU-filene til et datasett med årstall som variabel, før flettingen mot stammen. Slik det gikk nå gjennomførte datasteget en IF-test for hvert datasett for hver observasjon, og det er lite hensiktsmessig når treff forekommer relativt sjelden.

### 6.c. Innlesning av IGU-filer.

Programmet (E.3. IGUFIL) leser inn en og en årgang fil med igangværende utdanninger, plukker de elevene og studentene som finnes i hovedutvalget, og beholder siste observasjon for hvert individ. Det blir da noe vilkårlig hvilken utdanning vi får med for de som er i gang med flere utdanninger, men dette

er trolig et lite problem. På noen av de nyere årgangene viste det seg nødvendig å sortere etter fødselsnummer, og her ble det også sortert etter klassetrinn slik at bare høyeste utdanningsaktivitet kom med. Igangværende utdanning ble tatt med på tilsvarende måte, men etter gammel utdanningsstandard med 5-sifret utdanningens art.

#### 6.d. Utdanningsdata II.

Programmet (E.4. UTD2DATA) fletter ("merge") på tilsvarende måte "stammen" (se kapittel 8) med de utdragene av årgangene med IGU-filer. Også her legges til dummyvariabelen "student", men her betyr '1' at personen er under utdanning (programmeringsfeil) og '2' at personen ikke er funnet på IGU-filen. Feilen er rettet opp på den krypterte versjonen, slik at '2' betyr at personen er under utdanning og '1' ikke gjenfunnet på IGU-filen. Programmet brukte urimelig mye CPU-tid (107 minutter), og dette skyldes tilsvarende problemer som med BHU-filene. Eneste forskjell var at dette omfattet 11 årganger mot 5 årganger, slik at problemet ble mer påtagelig. Her burde spesielt IGU-filene vært flettet på forhånd med årstall som variabel, fordi bare 10 prosent av befolkningen er under utdanning hvert år.

## **7. Trygde- og yrkeshistorier.**

Siktemålet med dette prosjektet har vært å få med kjennetegn knyttet til alders- og uføretrygd fra Folketrygden i modellen MOSART. I utgangspopulasjonen har det vært viktig å kjenne trygdestatus og i enda større grad hittil opptjente pensjonsrettigheter. I tillegg ønsker vi å analysere overganger til å motta trygd, og det har da vært ønskelig med litt mer detaljerte opplysninger om trygdestatus, enn det som ligger i opptjeningen av pensjonsrettigheter. Fra registeret for pensjonpoeng i Rikstrygdeverket har vi hentet ut pensjonspoeng, pensjongivende inntekt, beregnet uførpoeng og høyeste uføregrad i løpet av året. Skattekommune skal også finnes i registeret, men dette var ikke teknisk lett tilgjengelig, og ble derfor utelatt. Fra filene med trygdemottakere (alders-, uføre- og etterlattetrygd; GR1) har vi hentet inn trygdestatus og trygdegrad ved utgangen av året. Trygdegrad vil for uføretrygdede si uføregrad og for alderspensjonister hvor stor del av trygden de tar ut. Det er klare inkonsistenser mellom pensjonspoengrekkene og GR1, i den grad at uførestatus dukker opp på ett senere tidspunkt i GR1. I forhold til offisiell statistikk er det GR1 som helt klart treffer best, og anbefales brukt ved analyser med mere. Opplysninger om trygdeytelser burde vært tatt med fra GR1. Spesielt er beregning av ventetillettet vanskelig/umulig bare på grunnlag av poengrekkene.

### **7.a. Data fra Rikstrygdeverket.**

Uttak av data fra Rikstrygdeverket må avtales nærmere med Rikstrygdeverket, av hensyn til arbeidsbelastning og hensiktsmessig utføring av arbeidet. Programmene (F.1a. UTVDISK/F.1b. UVTAPE) plukker ut de deler av utvalget som er eldre enn femten år, og legger dette over på magnetbånd som kan sendes Rikstrygdeverket. Individer som forekommer flere ganger blir bare representert med en observasjon. I retur fra Rikstrygdeverket fikk vi en sekvensiell fil med en observasjon for hver person for hvert år denne finnes i pensjonspoengregistret. Denne filen overføres til stormaskinen av driftskontoret, og programmet (F.2. UTVDISK) fletter ("merge") disse dataene mot stammen (se kapittel 8). Variable som legges inn er pensjongivende inntekt i løpende kroner, pensjonspoeng, beregnet uførpoeng og høyeste uføregrad i løpet av året.

### **7.b. Innlesning av GR1-data.**

Programmet (F.4. GR1DATA) leser inn en og en GR1-fil (mottakere av trygd fra Folketrygden) og legger de som tilhører hovedutvalget ut på disk. Det er av den grunn et vesentlig poeng å holde et visst tempo i denne delen av arbeidet. Variable som leses inn er årstall, indikator for å være trygdemottaker ved utgangen av året, trygdestatus og grad av uførhet og alderstrygd.

### **7.c. GR1-data.**

Programmet (F.5. TRYDATA) fletter ("merge") først sammen de ulike årgangene GR1-filer etter fødselsnummer og årstall, før blokken GR1-data fletter mot stammen (se kapittel 8). Dette brakte forbruket av CPU-tid



betraktelig nedover, jamfør avsnitt (6.b) og (6.d) om utdanning. Variable som beholdes på resultatfilene er trygdestatus og trygdegrad som angir grad av eventuell uførhet eller alderspensjon.

## 8. Tilrettelegging av data.

Dataene er i hovedsak tilrettelagt gjennom fire filer:

- Hovedutvalget.
- Ekteskapsmeldinger.
- Fødselsmeldinger.
- Individhistoriefil.

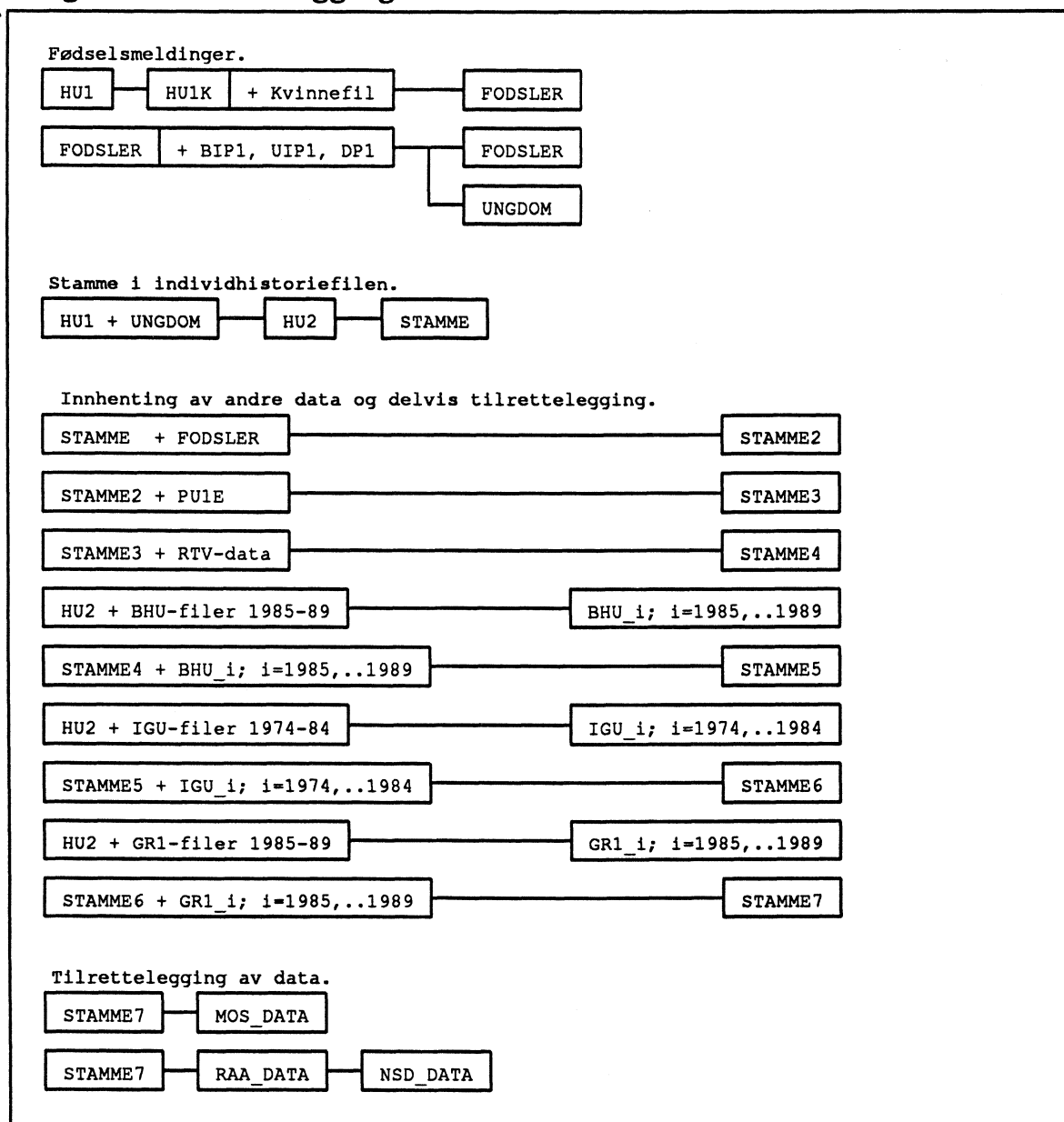
De tre første er i stor grad lik de filer som tidligere har blitt omtalt under kapittel (2) og (4). Til hovedutvalget har vi lagt alle de ungdommene under 16 år som skal brukes ved endogen rekruttering, mens fødselsmeldingene omfatter alle de fødslene som er lest inn. Hovedresultatet er imidlertid individhistorie-filen hvor hvert individ er representert med en observasjon for hvert år dette individet kan ha vært bosatt i Norge. Ved forløpsanalyse med diskret tid er dette en behagelig lagringsform. Hver observasjon inneholder statusvariable for gjeldende år, samt noen støttevariable som er like fra år til år, men som letter bruken av filen. En viss grense er satt, for eksempel er bare alder for yngste barn tatt med, og alder for eventuelle eldre barn finnes bare på fødselsmeldingsfilen. Individhistoriefilen er bygd opp ved å bygge opp en stamme, for deretter å legge lag på lag med data på denne. Stammen omfatter fødselsnummer og noen andre viktige opplysninger fra hovedutvalget og genererte årstall. Hvert individ er av EDB-tekniske årsaker bare representert med en tidsserie. Individuer som forekommer flere ganger er markert med egen indikator for dette. På stammen har vi lagt på i rekkefølge fødselshistorier (kvinner), ekteskaphistorier, pensjonspoengrekker, utdanningshistorier og trygdehistorier. En sekvensiell oppbygging er valgt for å beholde oversikten (kontroll) i koblingsarbeidet.

Disse fire resultatfilene er oppbevart på to forskjellige måter, ett sett filer for bruk i MOSART og ett sett filer for eventuelle andre brukere. Et viktig poeng ved (langtids)lagring er hensynet til personvernet og en tilstrekkelig anonymisering av filene. For bruk i modellen MOSART har vi lagt ekstra vekt på dette, da vi simulerer de videre livsløpene med følsomme begivenheter som dødelighet, skilsmisse med mere. En bakveisidentifikasjon her må ikke være mulig. På MOSART-filene er derfor fødselsnummer erstattet med et vilkårlig id.nummer, datoer gjennomgående erstatter med årstall, utdanning aggregert og bostedskommune fjernet. MOSART-filene ligger som SAS-filer, med back-up på stormaskinen og filbeskrivelse blant annet i vedlegg (B). I tillegg er de fire resultatfilene overført til sekvensielle filer (systemuavhengige) for å være mer generelt tilgjengelige. På disse filene er fødselsnummeret "kryptert" og datoer er gjennomgående erstattet med årstall. Disse filene har fått en begrenset lagringstid på fem år. Figur (2) på neste side viser gangen i tilretteleggingen av data.

### 8.a. Hovedutvalget.

Programmet (G.1. HU2DATA) plukker ut de ungdommene under 16 år fra

**Figur 2. Tilrettelegging av data.**



filen for fødselsmeldinger som er bosatt i Norge pr 1/1-1990, og slår disse sammen med hovedutvalget. Til slutt sorteres utvalget på fødselsnummer og utvalgsstatus. Dette utvalget inngår som en av de fire resultatfilene, og utgjør kjernen ved opprettelsen av stammen. Ungdommer som er døde eller utvandret er inkludert gjennom trekningen av ikke bosatte.

### 8.b. Oppretting av stammen.

Programmet (G.2. STAMME) tar hovedutvalget fra forrige avsnitt, og omgjør hvert individ til en observasjon med opplysninger om reg.status og visse andre statusvariable pr 1/1-1990. Disse individene blåses deretter opp slik at de får en observasjon for hvert år de kan ha vært bosatt i Norge utfra de data vi har.

Til hver observasjon legges indikator for tilknytning til utvalget (dublett) og antatt reg.status det året (vital). Her antar vi at innvandrere ikke har bodd i Norge før og at personer som er døde eller har utvandret, i sin helhet bodde i Norge før denne begivenheten. De andre dataene legges til stammen, se detaljkapitler om de ulike delene av livshistoriene.

### **8.c. Anonymisering av MOSART-filer.**

De filene som skal brukes i MOSART-T er de resultatfilene som har framkommet ved koblingen, men hvor fødselsnummer er anonymisert i tillegg til noen andre forandringer. Programmet (G.3. ANONYM) tar hovedutvalget fra (8.a), overført til disk og fjernet for flere observasjoner pr individ, og tildeler hvert fødselsnummer et tilfeldig tall. Deretter sorteres utvalget på denne tilfeldige variabelen, og hvert individ tildeles et id.nummer lik observasjonsnummeret. Vi er da sikret at dette id.nummeret er en-entydig, tilfeldig og senere ikke mulig å rekonstruere. Til slutt sorteres filen på fødselsnummer for å kunne påføres resultatfilene.

Programmet (G.4. ANOHU) legger id.nummeret til hovedutvalget ved å flette ("merge") på fødselsnummer, deretter omgjøres datoer til årstall og alle fødselsnumre slettes. Til slutt sorteres resultatfilen på id.nummeret. Programmet (G.5. ANOPU) legger først id.nummeret til hovedpersonen på ekteskapsmeldingene. Deretter sorteres det på ektefelles fødselsnummer, og id.nummeret legges til ektefelles id.nummer. Til slutt slettes alle fødselsnumre og datoer omgjøres til årstall, og resultatfilen sorteres på hovedpersonens id.nummer. Programmet (G.6. ANOFM) legger først id.nummeret til morens fødselsnummer på fødselsmeldingene. Deretter sorteres filen på barnets fødselsnummer, og tilsvarende gjøres for barnets fødselsnummer. Tilslutt slettes alle fødselsnumre og datoer omgjøres til årstall, og resultatfilen sorters på morens fødselsnummer. Programmet (G.7. ANOSTAM) legger id.nummeret til individhistoriefilen. Samtidig omgjøres alle datoer til årstall, utdanning aggregeres, inntekter gjøres om til realverdier (1989-kroner) og bostedskommune og fødselsnumre slettes. Da filen neppe lar seg sortere med rimelig innsats, blir ikke ektefelles fødselsnummer erstattet av id.nummer. Filen er av samme grunn ikke sortert kronologisk på id.nummer, men ved å bruke "NOTSORTED" kan dataene brukes som at de er gruppert etter id.nummer.

### **8.d. Sekvensielle filer.**

For at filene skal kunne krypteres må de overføres til sekvensiell filer (ASCII) med fast recordlengde. Dette er en enkel operasjon i SAS, og programmene (H.1. RAHU/H.2. RÆKT/H3. RAFODS/H.4a. RASTAM) overfører resultatfilene. Merk at "stammen" må deles i to, da rutingen p.t. for erstatning av fødselsnummer med såkalt S-nummer ikke klarer å håndtere så store datamengder. I tillegg har vi valgt å erstatte datoer med årstall, for å redusere muligheten for "bakveisidentifikasjon" (fødselsdager m.m.).

### **8.e. Kryptering.**

Første del av krypteringen består i å erstatte alle fødselsnummer med et såkalt et statistikknummer (S-nr). Denne operasjonen kan gjøres fra rutinen

"krypton" under "adabas/natural-produksjon". Etter at fødselsnummeret er byttet ut, slås stammen sammen til en fil og dette gjøres i programmet (H.4b. RASTAM2). Veiledning på bruk av rutinen KRYPTON og ansvaret for den videre krypteringen ligger p.t. under seksjon for EDB. Rette person her vil ordne kryptering av filene, som består av å erstatte S-nummer med et kryptert nummer (K-nr). Dette K-nummeret gir for utenforstående ingen annen informasjon enn kjønn og alder (hele måneder). Derimot kan de som sitter med kodenøkkelen fortsatt koble til ytterligere informasjon. Filene som kommer ut av krypteringen er kun sortert på K-nummer. Det er derfor nødvendig å sortere stammen på K-nummer og årstall for data, og dette gjøres i programmet (H.4c. RASTAMS). Videre tas det back-up kopier av filene. SAS bruker tegnet '' som blank/uoppgitt for numeriske verdier, og dette har gitt noen mindre problemer i krypteringen.



# **Vedlegg.**

## A. Filbeskrivelser, Mellomresultater.

### Hovedutvalget.

Datatype : Sas-data.  
Filident : S6320.S6921.I855C6A1.G8900.V03/hul.  
Opprettet: 26.6.91.  
Obs.enhet: Person, en observasjon hver gang personen er trukket ut.  
Sortering: Fnr\*utv\_stat(descending).  
Observasjoner: 477.434.  
Variable (10 - 35 Byte):

fnr	- fødselsnummer . . . . .	(6).
fdato	- fødselsdato . . . . .	(date9).
kjonn	- kjønn . . . . .	(Å1).
s_reg	- registreringsstatus . . . . .	(Å1).
s_rd	- dato for registreringsstatus . . . . .	(date9).
s_ekt	- ekteskapeleg status . . . . .	(Å1).
s_ed	- dato for ekteskapeleg status . . . . .	(date9).
e_fnr	- eventuell (siste) ektefelles fødselsnummer . . . . .	(6).
del_utv	- hvilket delutvalg tilhører personen . . . . .	(2).
utv_stat	- hvilken tilknytning har personen til utvalget . . . . .	(Å1).

### Ekteskapsmeldinger.

Datatype : Sas-data.  
Filident : S6320.S6921.I885C6A2.G8900.V03/pule.  
Opprettet: 27.6.91.  
Obs.enhet: Endring i ekteskapeleg status for person.  
Sortering: Fnr\*dato.  
Observasjoner: 360.725.  
Variable (8 - 33 Byte):

fnr	- fødselsnummer . . . . .	(6).
fdato	- fødselsdato . . . . .	(date9.).
kjonn	- kjønn . . . . .	(Å1).
prior	- Prioritet ved inkonsistente meldinger . . . . .	(Å1).
type	- type melding . . . . .	(Å1).
e_dato	- dato for ekteskapsmeldingen . . . . .	(date9).
e_aar	- årstall for ekteskapsmeldingen . . . . .	(3).
e_fnr	- ektefelles fødselsnummer . . . . .	(6).



## Ungdom.

Datatype : Sas-data.  
Filident : S6320.S6921.I885C8A2.G8900.V01/ungdom.  
Opprettet: 3.7.91.  
Obs.enhet: Fødsel/barn.  
Sortering: B\_fnr.  
Observasjoner: 89.476.  
Variable (11 - 40 Byte):

mor_fnr	- mors fødselsnummer	(6).
del_utv	- mors delutvalg	(Å1).
far_fnr	- fars fødselsnummer	(6).
treff_i	- er barnet gjenfunnet på statusfilen	(Å1).
b_fnr	- barnets fødselsnummer	(6).
b_fdato	- fødselsdato	(date9.).
b_f_aar	- året barnet er født	(3).
b_kjonn	- kjønn	(Å1).
s_b_reg	- barnets reg.status	(Å1).
s_b_red	- dato for barnets reg.status	(6).
s_b_ra	- årstall for barnets reg.status	(3).

## Fødselsmeldinger.

Datatype : Sas-data.  
Filident : S6320.S6921.I885C8A3.G8900.V01/fodsler.  
Opprettet: 3.7.91.  
Obs.enhet: Fødsel/barn.  
Sortering: M\_fnr\*b\_fdato.  
Observasjoner: 262.161.  
Variable (11 - 40 Byte):

mor_fnr	- mors fødselsnummer	(6).
del_utv	- mors delutvalg	(Å1).
far_fnr	- fars fødselsnummer	(6).
treff_i	- er barnet gjenfunnet på statusfilen	(Å1).
b_fnr	- barnets fødselsnummer	(6).
b_fdato	- fødselsdato	(date9.).
b_f_aar	- året barnet er født	(3).
b_kjonn	- kjønn	(Å1).
s_b_reg	- barnets reg.status	(Å1).
s_b_red	- dato for barnets reg.status	(6).
s_b_ra	- årstall for barnets reg.status	(3).

## Stamme7.

Datatype : Sas-data.  
Filident : S6320.S6921.I885D1A7.G8900.V01/stamme7.  
Opprettet: 23.9.1990.  
Obs.enhet: Person \* år.  
Sortering: Fnr \* hist\_aar.  
Observasjoner: 10.059.952.  
Variable: Kjerne (2 - 9 Byte).  
Totalt (26 - (84 Byte)).

Kjerne (2 variable - 9 byte).

fnr - fødselsnummer . . . . . (6).  
hist\_aar - årstall for observasjonen . . . . . (3).

Stamme (6 variable - 12 byte).

del\_utv - hvilket delutvalg tilhører personen . . . . . (2).  
dublett - hvilken tilknytning har personen til utvalget . . . . . (Å1).  
fdato - fødselsdato . . . . . (6).  
kjonn - kjønn . . . . . (Å1).  
s\_ekt - ekteskapeleg status pr 1/1-1990 . . . . . (Å1).  
vital - vital definert ved reg.status pr 1/1-1990. . . . . (Å1).

Ekteskapshistorier (3 variable - 8 byte).

ekt\_stat - ekteskapeleg status . . . . . (Å1).  
ekt\_fnr - ektefelles fødselsnummer . . . . . (6).  
e\_ekt - endring i ekteskapeleg status . . . . . (Å1).

Fodselslshistorier (bare kvinner) (2 variable - 4 byte).

ant\_barn - antall barn . . . . . (2).  
ald\_yb - yngste barn's alder . . . . . (2).

Utdanningshistorier (7 variable - 25 byte).

hfu\_art - høyeste fullførte utdanning, art, ny kode. . . . . (Å6).  
hfu\_klt - høyeste fullførte utdanning, klassetrinn, ny kode. . . . . (Å2).  
hfu\_ff - høyeste fullførte utdanning, fullføringstidspunkt . . . . . (Å4).  
igu\_art - igangværende utdanning, art, ny kode . . . . . (Å6).  
igu\_klt - igangværende utdanning, klassetrinn, ny kode . . . . . (Å2).  
student - dummy for utdanningsaktivitet . . . . . (Å1).  
kommune - bostedskommune . . . . . (Å4).

Pensjonspoenghistorier (4 variable - 13 byte).

pp - pensjonpoeng . . . . . (5).  
inntekt - pensjongivende inntekt . . . . . (3).  
bup - beregnet uførepoeng . . . . . (3).  
ufo\_grad - høyeste uføregrad i løpet av året . . . . . (2).

Trygdehistorier (2 variable - 3 byte).

try\_stat - trygdestatus . . . . . (Å1).  
try\_grad - trygdegrad (for alders- og uførepensjonister) . . . . . (2).

## B. Filbeskrivelser, MOSART.

### Hovedutvalget.

Datatype : Sas-data.  
Filident : SL320.S6921.I855E1A1.G6789.V00/hu2.  
Opprettet: 14.10.91.  
Obs.enhet: Person, en observasjon hver gang personen er trukket ut.  
Sortering: Idnr\*utv\_stat(descending).  
Observasjoner: 563.963.  
Variable (9 - 18 Byte):

idnr	- anonymisert fødselsnummer . . . . .	(4).
f_aar	- fødselsår . . . . .	(3).
kjonn	- kjønn . . . . .	(Å1).
s_reg	- registreringsstatus . . . . .	(Å1).
s_r_aar	- år for registreringsstatus . . . . .	(3).
s_ekt	- ekteskkelig status . . . . .	(Å1).
s_e_aar	- dato for ekteskkelig status . . . . .	(3).
del_utv	- hvilket delutvalg tilhører personen . . . . .	(2).
utv_stat	- hvilken tilknytning har personen til utvalget . . . . .	(Å1).

### Ekteskapsmeldinger.

Datatype : Sas-data.  
Filident : SL320.S6921.I855E1A2.G6789.V00/pule2.  
Opprettet: 14.10.91.  
Obs.enhet: Endring i ekteskkelig status for person.  
Sortering: Idnr\*år.  
Observasjoner: 360.725.  
Variable (5 - 13 Byte):

idnr	- anonymisert fødselsnummer . . . . .	(4).
prior	- Prioritet ved inkonsistente meldinger . . . . .	(Å1).
type	- type melding . . . . .	(Å1).
e_aar	- årstall for ekteskapsmeldingen . . . . .	(3).
e_idnr	- ektefelles idnr . . . . .	(4).

### Fødselsmeldinger.

Datatype : Sas-data.  
Filident : SL320.S6921.I885E1A3.G6789.V00/fodsler.  
Opprettet: 14.10.91.  
Obs.enhet: Fødsel/barn.  
Sortering: Mor\_idnr\*b\_f\_aar.  
Observasjoner: 262.161.  
Variable (8 - 18 Byte):

mor_idnr	- mors idnr . . . . .	(4).
del_utv	- mors delutvalg . . . . .	(Å1).
treff_i	- er barnet gjenfunnet på statusfilen . . . . .	(Å1).
b_idnr	- barnets idnr . . . . .	(4).
b_f_aar	- året barnet er født . . . . .	(3).
b_kjonn	- kjønn . . . . .	(Å1).
s_b_reg	- barnets reg.status . . . . .	(Å1).
s_b_ra	- årstall for barnets reg.status . . . . .	(3).

## Stamme/individhistoriefil.

Datatype : Sas-data.  
Filident : SL320.S6921.I885E1A4.G6789.V00/mosdata.  
Opprettet: 15.10.91.  
Obs.enhet: Person \* år.  
Sortering: Gruppert etter idnr \* hist\_aar.  
Observasjoner: 10.059.952.  
Variable: Kjerne (2 - 9 Byte).  
          Totalt (23 - 52 Byte).

### Kjerne (2 variable - 7 byte).

idnr - anonymisert fødselsnummer . . . . . (4).  
hist\_aar - årstall for observasjonen . . . . . (3).

### Stamme (6 variable - 11 byte).

kjonn - kjønn . . . . . (Å1).  
f\_aar - fødselsår . . . . . (3).  
alder - alder . . . . . (2).  
del\_utv - hvilket delutvalg tilhører personen . . . . . (2).  
dublett - hvilken tilknytning har personen til utvalget . . . . . (Å1).  
s\_ekt - ekteskadelig status pr 1/1-1990 . . . . . (Å1).  
vital - vital definert ved reg.status pr 1/1-1990. . . . . (Å1).

### Ekteskaphistorier (2 variable - 2 byte).

ekt\_stat - ekteskadelig status . . . . . (Å1).  
e\_ekt - endring i ekteskadelig status . . . . . (Å1).

### Fodselskshistorier (bare kvinner) (2 variable - 4 byte).

ant\_barn - antall barn . . . . . (2).  
ald\_yb - yngste barn's alder . . . . . (2).

### Utdanningshistorier (5 variable - 12 byte).

hfu\_fag - høyeste fullførte utdanning, art, aggregert . . . . . (Å2).  
hfu\_klt - høyeste fullførte utdanning, klassetrinn, aggregert . . . . . (Å2).  
hfu\_ff - høyeste fullførte utdanning, fullføringstidspunkt . . . . . (Å4).  
igu\_art - igangvarende utdanning, art, ny kode . . . . . (Å2).  
igu\_klt - igangvarende utdanning, klassetrinn, ny kode . . . . . (Å2).

### Pensjonspoenghistorier (4 variable - 13 byte).

pp - pensjonpoeng . . . . . (5).  
inntekt - pensjongivende inntekt, deflatert . . . . . (3).  
bup - beregnet uførepoeng . . . . . (3).  
ufo\_grad - høyeste uføregrad i løpet av året . . . . . (2).

### Trygdehistorier (2 variable - 3 byte).

try\_stat - trygdestatus . . . . . (Å1).  
try\_grad - trygdegrad (for alders- og uførepensjonister) . . . . . (2).

## C. Filbeskrivelser, sekvensielle filer.

### Hovedutvalget.

Datatype : Sekvensiell fil.  
Filident : PL320.S6921.I855F1A1.G6789.V00K.  
Opprettet: Mars 1992.  
Obs.enhet: Person, en observasjon hver gang personen er trukket ut.  
Sortering: Kryptert fødselsnummer.  
Observasjoner: 477.434.  
Recordbeskrivelse (34 byte):

1 - 11 Kryptert fødselsnummer (se vedlegg D. formater).  
12 Utv\_stat  
13 Del\_utv  
14 Regstatus (s\_reg)  
15 - 18 Årstall fra reg.status (fra s\_rd)  
19 Ekteskapelig status (s\_ekt)  
20 - 23 Årstall for ekteskapelig status (fra s\_ed)  
24 - 34 Ektefelles krypterte fødselsnummer

### Ekteskapsmeldinger.

Datatype : Sekvensiell fil.  
Filident : PL320.S6921.I885F1A2.G6789.V00K.  
Opprettet: Mars 1992.  
Obs.enhet: Endring i ekteskapelig status for person.  
Sortering: Kryptert fødselsnummer.  
Observasjoner: 360.725.  
Recordbeskrivelse (28 byte):

1 - 11 Kryptert fødselsnummer (se vedlegg D. formater).  
12 Prior  
13 Type melding (type)  
14 - 17 Årstall for endret ekt.status (s\_r\_aar)  
18 - 28 Ektefelles krypterte fødselsnummer.

### Fødselsmeldinger/ungdom.

Datatype : Sekvensiell fil.  
Filident : PL320.S6921.I885F1A3.G6789.V00K.  
Opprettet: Mars 1992.  
Obs.enhet: Fødsel/barn.  
Sortering: Mors krypterte fødselsnummer.  
Observasjoner: 262.161.  
Recordbeskrivelse (44 byte):

1 - 11 Mors krypterte fødselsnummer (se vedlegg D. formater).  
12 Mors delutvalg (del\_utv)  
13 - 23 Fars krypterte fødselsnummer  
24 - 34 Barnets krypterte fødselsnummer  
35 - 38 Barnets fødselsår (firesifret)  
39 Treff\_i  
40 Barnets reg.status (s\_b\_reg)  
41 - 44 År for barnets reg.status (s\_b\_ra)

## Stamme7.

Datatype : Sekvensiell fil.  
Filident : PL320.S6921.I885F1A4.G6789.V00K.  
Opprettet: Mars 1992.  
Obs.enhet: Person \* år.  
Observasjoner: 10.059.952.  
Sortering: Kryptert fødselsnummer \* hist\_aar.  
Recordbeskrivelse (89 byte):

1 - 11 Kryptert fødselsnummer (se vedlegg D. formater).  
12 - 15 Årstall for data (hist\_aar)

17 Dublett  
18 Vital  
19 Del\_utv  
20 Ekteskapeleg status pr 1/1-1990 (s\_ekt)

22 Ekteskapeleg status fortløpende (ekt\_stat)  
23 - 33 Ektefelles krypterte fødselsnummer  
34 Endret ekteskapeleg status (e\_ekt)  
35 - 36 Antall barn (ant\_barn)  
37 - 38 Alder på yngste barn (ald\_yb)

40 - 45 Hfu\_art  
46 - 47 Hfu\_klt  
48 - 51 Fullføringsår for hfu  
52 - 57 Igu\_art  
58 - 59 Igu\_klt  
60 Student/elev  
61 - 64 Formell bostedskommune

66 - 69 Pensjonspoeng (pp, format punktum, 2 desimaler)  
70 - 78 Pensjongivende inntekt (inntekt, format heltall)  
79 - 82 Beregnet uførepoeng (bup, format punktum, 2 desimaler)  
83 - 85 Høyeste uføregrad (ufo\_grad, format heltall)  
86 Trygdestatus (try\_stat)  
87 - 89 Trygdegrad (try\_grad, format heltall)

## **D. Formater.**

Avsnittet gir en oversikt over de 'formater' eller forkortelser som er brukt for ulike verdier av viktige variable.

### **Kjonn.**

K=kvinner.  
M=menn.

### **Reg.status (\_reg).**

1=bosatt.  
2=død.  
3=utvandret.

### **Ekteskapeelig status (\_ekt).**

0=ukjent.  
1=ugift.  
2=gift.  
3=enke/enkemann.  
4=skilt.  
5=separert.

### **Dublett (tilknytning til hovedutvalget).**

1=primærutvalget.  
2=primærutvalget og tilknyttet person.  
3=tilknyttet person.

### **Utv\_stat (Utvalgsstatus).**

1=primærutvalget.  
2=avdød ektefelle.

### **Prioritet (ekteskapsmeldinger).**

1=statusfilen.  
2=andre.

### **Type (ekteskapsmeldinger).**

1:ugift (=fødsel).  
2:giftetål.  
3:ektefelles dødsfall.  
4:eget dødsfall.  
5:separasjon.  
6:skilsmisse.  
7:ukjent.

### **Treff\_i (fødselsmeldinger/ungdom).**

- 0: Ikke treff.
- 1: Barnet gjenfunnet på bipl, dpl eller uipl.

### **Vital (utvidet reg.status).**

- 1: Bosatt.
- 2: Død.
- 3: Utvandret.
- 4: Innvandrer (neste år).
- 5: Kommende innvandrer.
- 6: Antatt bosatt (pre utvandring og død).

### **Try\_stat (trygdestatus).**

- 1: Ikke mottaker av stønader 3-7.
- 2: Ikke uføretrygdet.
- 3: Uføretrygdet.
- 4: Alderstrygdet.
- 5: Etterlatt ektefelle.
- 6: Etterlatt familiepleier.
- 7: Etterlatt barn.

### **Student.**

Årene 1974-1984:

- 1: Ikke under utdanning.
- 2: Elev eller student.

Årene 1985-1989:

- Blank: Ikke treff mot BHU.
- 1: Ikke under utdanning.
  - 2: Elev eller student.

### **Kryptert fødselsnummer.**

- 1       Kodenøkkel.
- 2 - 8   K-nummer.
- 9       Kjønn/århundre.
  - 1: Mann, 1800-1899.
  - 2: Mann, 1900-1999.
  - 3: Mann, 2000-2099.
  - 4: Kvinne, 1800-1899.
  - 5: Kvinne, 1900-1999.
  - 6: Kvinne, 2000-2099.
  - 7: Uoppgitt kjønn.
- 10 - 11 Fødselsår.

Kryptert nummer 'A34655309 ' er uoppgitt fødselsnummer.



**Konsumprisindeksen.**

År	Indeks
67	18.9
68	19.5
69	20.2
70	22.3
71	23.7
72	25.4
73	27.3
74	29.9
75	33.4
76	36.4
77	39.7
78	43.0
79	45.0
80	49.9
81	56.7
82	63.2
83	68.5
84	72.8
85	76.9
86	82.4
87	89.6
88	95.6
89	100.0
90	104.1

## **E. Annen dokumentasjon og litteratur.**

Andreassen, Leif og Dennis Fredriksen (1991): "MOSART - en mikrosimuleringsmodell for utdanning og arbeidsstyrke", Økonomiske analyser 91/2, Statistisk sentralbyrå.

Brunborg, Helge: "Kvalitetskontroll av ekteskapsfilen", Interne Notater 86/32, Statistisk sentralbyrå.

Fredriksen, Dennis (1991): "To notater fra prosjektet MOSART og trygd", Interne Notater 91/8, Statistisk Sentralbyrå.

Fredriksen, Dennis (1991b): "Vedlegg til publikasjoner på MOSART: Aggregering av utdanning.", upublisert notat.

Fredriksen, Dennis (1992): "Datagrunnlaget for trygdemodellen, databearbeidende programmer", upublisert notat.

Kravdal, Øystein: "Teknisk notat om oppbygging av mødrefil, ekteskapsfil og kvinnefil", Interne Notater 86/27, Statistisk sentralbyrå.

Standarder for Norsk statistikk: "Standard for utdanningsgruppering", Statistisk sentralbyrå 1989.

Vassenden, Elisabetta (1990): "Befolkningens høyeste fullførte utdanning, dokumentasjon", Interne Notater 90/9, Statistisk sentralbyrå.