

*Mona Takle*

**Befolkningsstatistikk på rutenett**  
Dokumentasjon

# Notater

# Innhold

<b>1. SAMMENDRAG .....</b>	<b>3</b>
<b>2. INNLEDNING .....</b>	<b>4</b>
2.1. BAKGRUNN .....	4
2.2. EKSEMPEL PÅ BRUK.....	6
<b>3. METODE .....</b>	<b>8</b>
3.1. KARTLEGGING AV TILSVARENDE PROSJEKT .....	8
3.2. ALTERNATIV 1 .....	8
3.3. ALTERNATIV 2 .....	8
3.4. ALTERNATIV 3 .....	8
3.5. KONKLUSJON .....	8
<b>4. DATAGRUNNLAG; KOORDINATFESTING AV BEFOLKNINGEN.....</b>	<b>10</b>
4.1. ADRESSEBEGREPET .....	10
4.2. KOBLING PÅ 17 POSISJONER .....	10
4.3. KOBLINGENE VIDERE .....	11
<b>5. TILRETTELEGGING AV DATA .....</b>	<b>13</b>
5.1. LAGE RUTE-ID PÅ BEFOLKNINGSDATAENE .....	13
5.2. TILRETTELEGGING AV BEFOLKNINGSTABELLER FOR ARCVIEW .....	13
5.3. LAGE RUTENETTET .....	15
5.4. PRODUKSJONSLINJE.....	17
<b>6. HENSynet TIL PERSONVERN.....</b>	<b>18</b>
6.1. PROBLEMSTILLING .....	18
6.2. UTLEVERING AV OPPLYSNINGER .....	18
6.3. OFFENTLIGGJØRING AV OPPLYSNINGER.....	18
6.4. ALDERSGRUPPERING .....	19
6.5. PRIKKING .....	20
6.6. AVRUNDING .....	21
6.7. FLERE VARIABLER?.....	21
6.8. STØRRELSE PÅ RUTENE .....	22
6.9. OPPSUMMERING .....	22
<b>7. NOEN RESULTATER.....</b>	<b>23</b>
7.1. FERDIG PRODUKT OG TESTBRUK.....	23
7.2. SAMMENLIGNING MED ANDRE NORDISKE LAND .....	24
<b>8. OPPSUMMERING .....</b>	<b>27</b>
8.1. MULIGHETER VIDERE FREMOVER .....	27
<b>Vedlegg</b>	
A Filstruktur .....	29
B Program for å lage rute-ID på befolkningsdataene.....	30
C Program for å lage rutenettet.....	32
De sist utgitte publikasjonene i serien Notater.....	34



# 1. Sammendrag

I dette notatet dokumenteres arbeidet med å lage befolkningsstatistikk på rutenett. Ettersom dette er et nytt produkt, har det vært en viktig del av prosjektet å skissere en produksjonslinje som beskriver hele prosessen. Notatet er beregnet på intern bruk.

I arbeidet med å lage rutenettstatistikk er det dratt veksler på det som er gjort i andre nordiske statistikkbyråer. Statistikkbyråene i Finland og Sverige har forholdsvis lang erfaring med rutenettstatistikk, Finland siden folketellingen i 1970. For tiden er også Danmarks Statistik, i samarbeid med Kort & Matrikelstyrelsen, i gang med et forprosjekt for etablering av statistikk på rutenett.

Statistikk på ruter bør tolkes og analyseres i et geografisk informasjonssystem (GIS). Det å tolke rutenettstatistikk i tabellform vil være vanskelig fordi rute-enheten først og fremst er meningsbærende når den knyttes til geografien. I et GIS vil rutenettet knyttes til geografien, og man kan kombinere statistikken med andre kartdata, som veger, vann, grøntarealer etc.

Produktet er svært fleksibel, både med hensyn til hvilke variabler som kan være med og i forhold til størrelsen på rutene. Ettersom det er et nytt produkt kan det imidlertid være fornuftig å tilby noen standardprodukter i en etableringsfase.

Det er enkelt og lite tidkrevende å produsere både rutenettet og statistikk på ruter. I første omgang legges det opp til årlig produksjon av befolkningsstatistikk på ruter. For hver rute vil det opplyses hvor mange som bor der, hvilket kjønn de har og hvilken aldersgruppe de tilhører:

rute-id	total	menn	kvinner	0-5 år	6-15 år	16-19 år	20-39 år	40-66 år	67 år +
---------	-------	------	---------	--------	---------	----------	----------	----------	---------

Visse justeringer med hensyn til variabler og størrelse på rutene bør vurderes etter hvert. På sikt bør SSB også presentere andre statistikkområder på rutenett.

Før offentliggjøring vil statistikken behandles slik at slik at den er i tråd med kravet til personvern.

## 2. Innledning

### 2.1. Bakgrunn

Administrative og statistiske inndelinger som fylker, kommuner og grunnkretser er av ulik størrelse og form. De kan inneholde store ubebodde områder eller områder der innbyggerne bor veldig tett. Grunnkretsene er pt. den minste regionale enheten det gis statistikk for i SSB, og altså den som gir det mest detaljerte bilde av bosettingsmønsteret. Selv om et av kriteriene ved etableringen av grunnkretssystemet var at grunnkretsene skulle være homogene (f.eks. når det gjelder befolkningsstruktur), er det slik at noen grunnkretser i dag inneholder både spredtbygde og tettbygde områder. Det kan derfor være vanskelig å danne seg et detaljert bilde av hvordan bosettingsmønsteret er.

SSB har i prinsippet muligheten til å tilordne en adressekoordinat for hver person som er registrert bosatt i Norge. All statistikk som kan relateres til fødselsnummer kan dermed i teorien koordinatfestes. Ved hjelp av GIS har man derfor muligheten til å aggregere statistikk til andre inndelinger enn de som tradisjonelt har blitt brukt for å presentere statistikk. Man kan for eksempel knytte statistikken til en ensartet inndeling, som et kvadratisk rutenett, slik at alle enhetene er like i utstrekning og form. Figur 2.1 viser fordelingen av bosettingen i Norge på rutenett.

Statistikk knyttet til et rutenett kan være nyttig som datagrunnlag i en rekke sammenhenger:

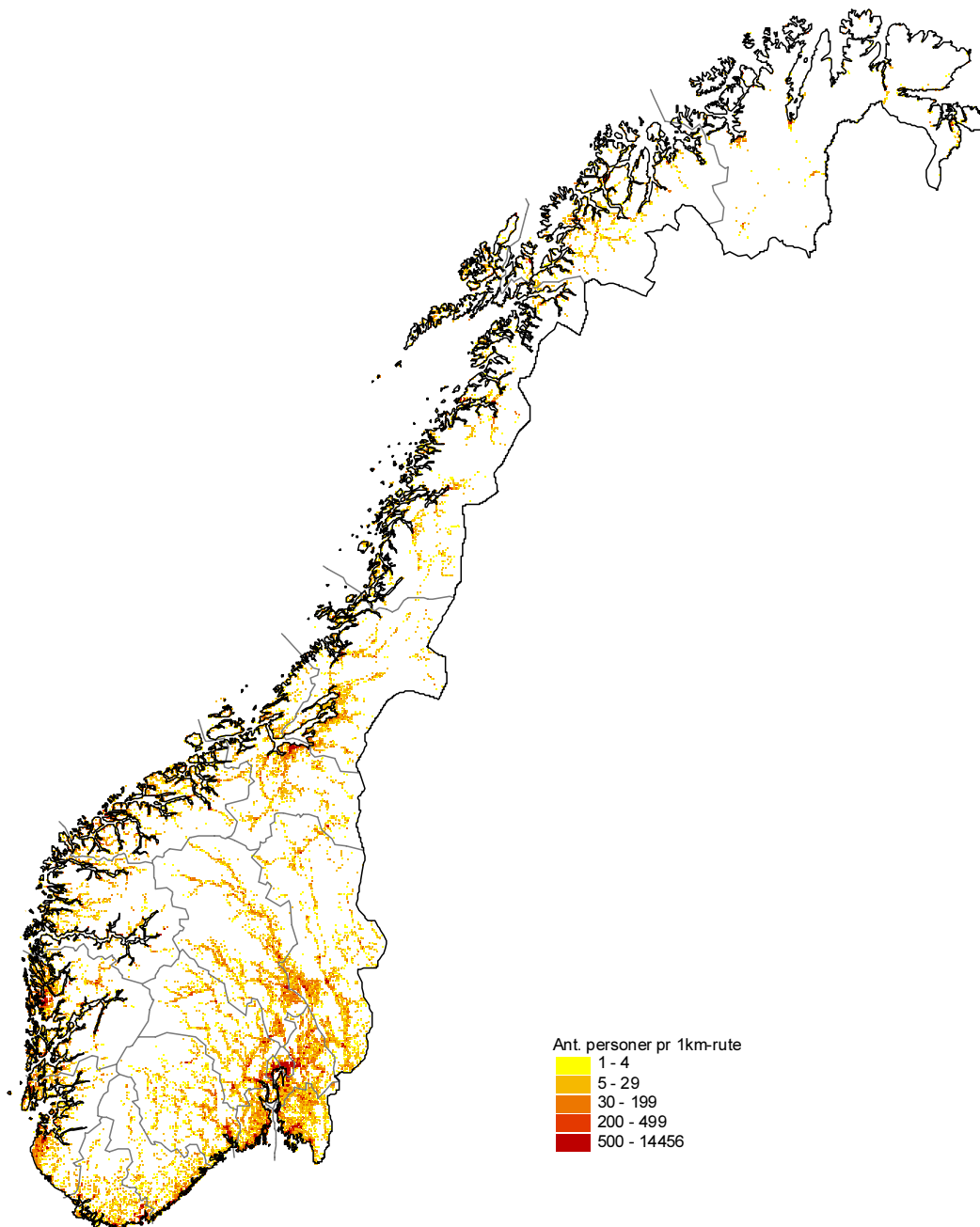
- det gir økt detaljeringsgrad i forhold til andre statistiske inndelinger.
- den regelmessige formen kan gi nye analysemuligheter, blant annet ved at det forenkler sammenligning mellom ulike geografiske områder. Når rutene får samme utstrekning over tid vil det også forenkles sammenligning mellom forskjellige årganger av statistikken (i forhold til andre inndelinger som gjerne endres over tid).
- rutene kan brukes som byggestener for å aggregere statistikken til andre, egendefinerte, inndelinger.

Av negative sider kan nevnes:

- rutenett-statistikk må sees i forbindelse med kart. Det er først og fremst når dataene knyttes til geografien de blir tolkbare. I geografiske informasjonssystemer er imidlertid analysemulighetene svært gode.
- rutene følger ikke geografiske eller sosio-økonomiske mønstre. Langs slike skillelinjer, f.eks. mellom land og sjø eller mellom to kommuner, vil verdien i ruten reflektere forekomstene på begge sider av skillelinjen. Veldig små ruter kan gi en oppløsning som i stor grad følger linjene, mens store ruter vil gi et unyansert bilde.

Rutene kan ikke erstatte andre typer statistikinndelinger som grunnkretser o.l., men de kan åpne for nye bruksmuligheter og være et bra tillegg til andre typer statistikk. Fra brukerne har SSB fått klare signaler om at det er behov for et slikt produkt.

Figur 2.1. Befolkningstetthet på kilometer-ruter i Norge. 1. januar 2000



## 2.2. Eksempel på bruk

Et rutenett-produkt vil være nyttig for flere brukergrupper. Blant annet kan nevnes offentlig forvaltning/planlegging, forskning, markedsføring, planlegging av private tjenester som telekommunikasjon, butikker, banker, byggeprosjekter etc.

I tillegg til å gi et generelt inntrykk av bosettingsmønsteret i Norge, kan produktet åpne opp for nye analysemuligheter. Noen eksempler kan være:

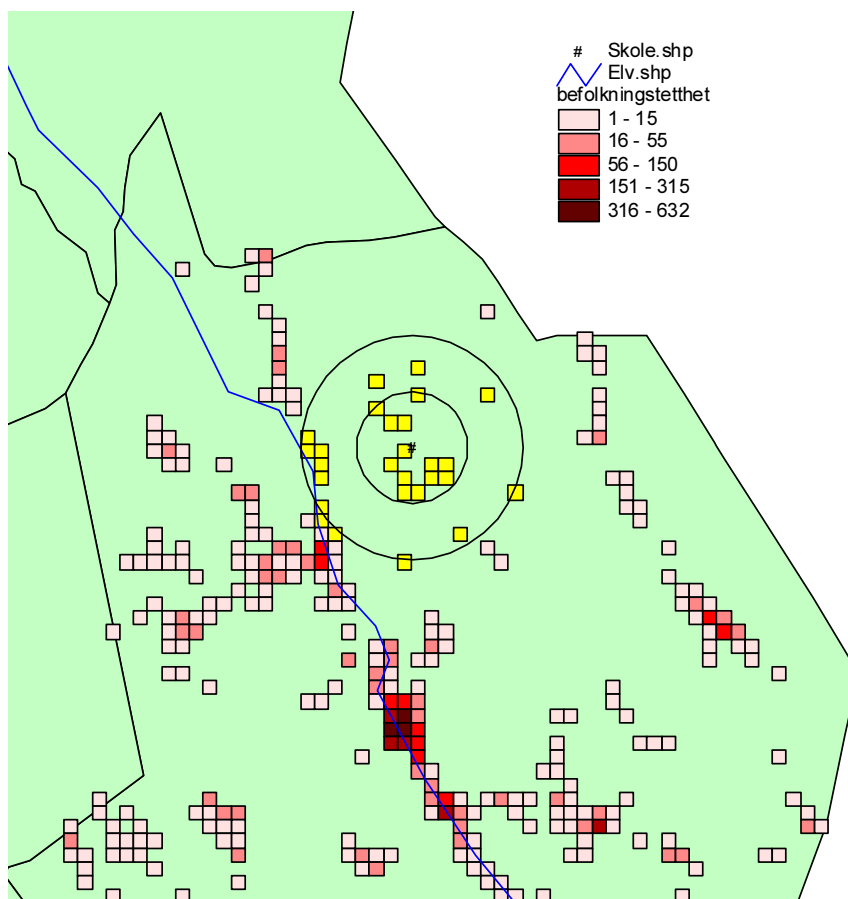
\* Hvor mange småbarn bor i en viss omkrets fra en skole? Fig. 2.2 viser en optelling av personer innenfor sonene 4000 m og 8000 m i aldersgruppen 0-5 år. I den innerste sirkelen bor det 5 og i begge sirkelene til sammen 19.

\* Hvor mange mennesker bor innen en viss buffersone fra en veg (forurensning, støy), busstrasé (tilgjengelighet) eller elv (oversvømmelse)? Figur 2.3 viser en optelling av personer innen en viss avstand fra en veg. På de valgte rutene bor det totalt 21 164 personer.

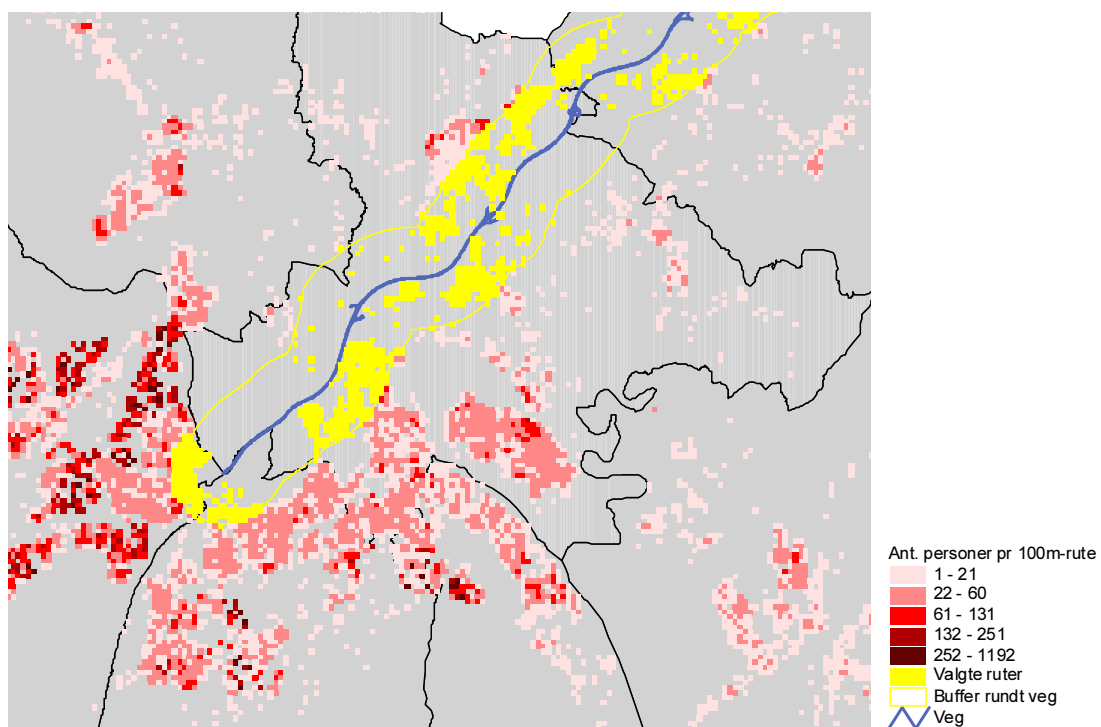
\* I hvilke områder bor det mange eldre? Figur 2.4 viser andelen eldre i forhold til resten av befolkningen. I de mørkeste rutene bor det flest eldre i forhold til totalt antall personer i ruten.

\* Etter hvert vil man også få muligheten til å se på endringer over tid ved å sammenligne datasett av forskjellig årgang. Man kan da for eksempel få frem befolkningsvekst/-nedgang i forskjellige områder.

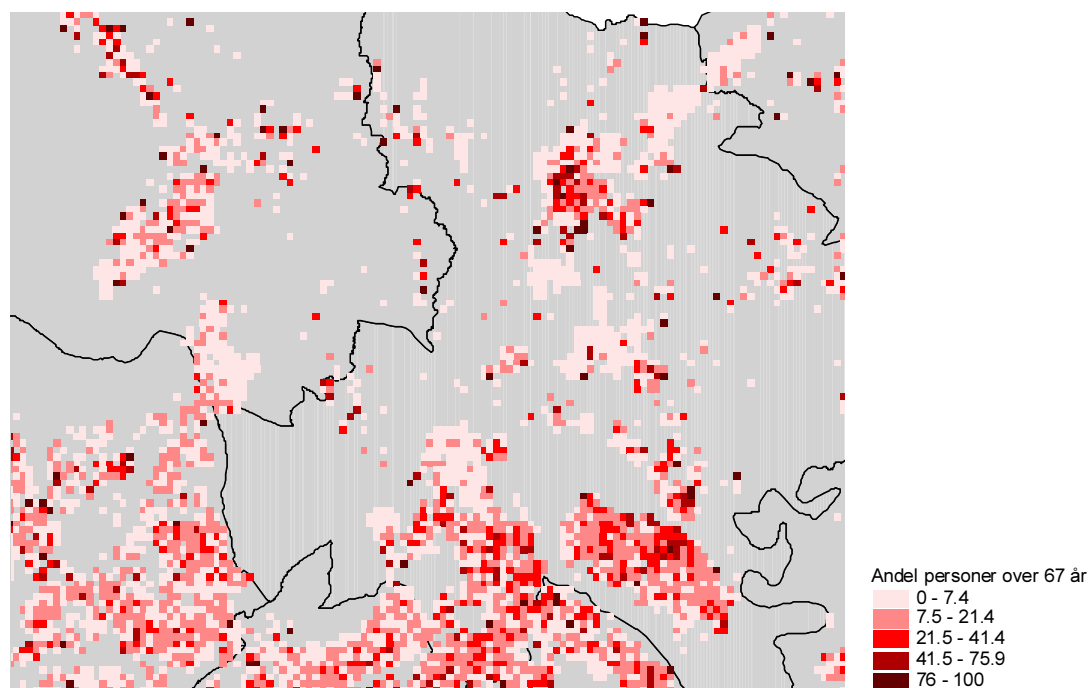
**Figur 2.2. Eksempel på søk i en viss radius rundt et punkt**



**Figur 2.3. Utvelging av ruter innen en viss buffersone fra en veg**



**Figur 2.4. Andel personer over 67 år i forhold til resten av befolkningen i ruten**





## 3. Metode

### 3.1. Kartlegging av tilsvarende prosjekt

Av de nordiske landene var Finland først ute med rutenettstatistikk, allerede i 1970. Det svenske statistikkbyrået (SCB) deltok i et samarbeidsprosjekt med Finland i 1995, der data på km-ruter fra de to landene bl.a. ble sammenlignet (Statistics Finland 1995). SSB leverte i 1999/2000 testdata som ble brukt til å fremstille et felleskart for befolkningen på rutenett for de tre landene. Danmarks Statistik (DS) er for tiden i ferd med å avslutte et forprosjekt for etablering av et "nationalt kvadratnett".

Også andre statistikkbyråer, bl.a. i Slovenia, Nederland og Israel, produserer statistikk på rutenett.

For å komme frem til den mest hensiktsmessige måten å lage rutenettstatistikk på for SSB ble forskjellige metoder vurdert. Her er en kort redegjørelse av disse.

### 3.2. Alternativ 1

Alternativ 1 går ut på å følge hovedtrekkene i rutenettarbeidet i Sverige og Finland. Produksjonen deles i to deler;

1. Man lager en statistikkfil med f.eks. befolkning pr. rute. Befolkningsregisteret inneholder opplysning om x/y-koordinat, og for hver person rundes koordinatene ned til nærmeste 1000 m eller annen aktuell rutenett-størrelse. De nye koordinatene slås sammen og brukes som en rute-ID.
2. Man etablerer et digitalt rutenettet. XY-koordinatene for hver rute (nedre, venstre hjørnet) blir slått sammen og brukt som en koblingsnøkkel/ID mot individdataene.

### 3.3. Alternativ 2

Alternativ 2 ble brukt ved leveransen av testdataene fra Norge til det nordiske rutenettprosjektet. Det går ut på å knytte befolkningsstatistikken til ruter i ArcView gjennom en "punkt-i-polygon"-prosess. Befolkningsstatistikken er da aggregert til adressekoordinater, og man teller opp hvor mange personer som bor i hver rute. Det mest negative ved denne fremgangsmåten er at den innebærer store datamengder og lang prosesseringstid.

### 3.4. Alternativ 3

Det siste alternativet som ble vurdert er å benytte raster programvare (f.eks. ESRI-produktet Spatial analyst). Rasterformat gir mye mindre datamengder enn et rutenett i vektorformat, men programvaren er ikke beregnet for å *aggregere* data til cellene. Samtidig begrenser dette formatet antallet brukere ettersom det ikke kan brukes av de som ikke har Spatial analyst.

Det kan imidlertid være en fordel å konvertere det endelige produktet til rasterformat. Datamengdene blir da atskillig mindre og man får nye analysemuligheter (f.eks. 3D-kart).

### 3.5. Konklusjon

I arbeidet med å legge en produksjonsløype for rutenett-statistikk i Norge, har SSB først og fremst dradd veksler på arbeidet som er gjort i de andre nordiske landene. Dette er naturlig i og med at

datagrunnlaget er forholdsvis likt. Dessuten vil en felles plattform forenkle sammenligning og samarbeid mellom landene. Dette taler for å velge alternativ 1.

Alternativ 1 er også den enkleste metoden når man skal behandle de store datamengdene det her er snakk om. Det vil da også være mulig å programmere en brukervennlig applikasjon for å hente ut statistikk med ønskede variabler og for ønsket geografisk område.

På bakgrunn av dette er det bestemt at alternativ 1 tas i bruk i dette prosjektet.

## 4. Datagrunnlag; koordinatfesting av befolkningen

### 4.1. Adressebegrepet

Prosjektet forutsetter at det finnes en adressekoordinat som representerer personers bosted. Koordinatfestet befolkning er viktig for SSB i flere sammenhenger, blant annet i forbindelse med tettstedsavgrensning, beregninger av reisetidsindikatorer i kommunene og statistikk på egendefinerte områder. Denne delen av prosjektet er altså felles for flere prosjekter.

Koordinatfestingen skjer i hovedsak ved en kobling mellom Statens kartverks register for Grunn-eiendommer, Adresser og Bygninger (GAB) og Det sentrale folkeregister (DSF). Koblingsnøkkelen mellom registrene er *numerisk adresse*, et begrep som spesifiseres nærmere under. Pga. manglende samstemthet mellom numerisk adresse i GAB og DSF, vil det være noen personer i DSF som ikke får påført koordinat. For noen av de som kobler vil det også mangle koordinat, fordi opplysningen mangler i GAB. Et annet kvalitetsproblem kan være at en adresse får påført koordinat, men at den er feil.

Det er altså tre potensielle feilkilder ved koordinatfesting av befolkningen:

1. DSF og GAB kobler ikke pga. ulik numerisk adresse
2. koordinaten mangler i GAB
3. koordinaten i GAB er feil

Med *adresse* menes en identifikasjon som entydig betegner et område/en bygning/bygningsdel innenfor en kommune. *Numerisk adresse* er en numerisk koding av adressebegrepene, og den skal i prinsippet skal være lik i DSF og GAB. En numerisk adresse kan være en gateadresse eller en matrikeladresse. Matrikeladressen brukes for områder som ikke har fått tildelt gateadresse, og den bygger på matrikelnummeret til en eiendom. En numerisk adresse består av forskjellige deler (nivåer). Disse beskrives i tabell 4.1.

**Tabell 4.1. Numerisk adresse**

Posisjon	Lengde	Gateadresse	Matrikeladresse
1 - 4	4	Kommunenummer	Kommunenummer
5 - 9	5	Gatekode	Gårdsnummer
10 - 13	4	Husnummer	Bruksnummer
14 - 17	4	Oppgang	Festenummer
18 - 20	3	Undernummer 2	Seksjonsnummer

### 4.2. Kobling på 17 posisjoner

SSB har koblet GAB og folkeregisteret fra og med 1990, blant annet for å koble grunnkretsopplysning fra GAB på hver person. Systematisk påkobling av koordinater har imidlertid vært gjort siden 1998. De første årene ble dette gjort via en kobling på 13 posisjoner av numerisk adresse; dvs. kommune-nummer (4 pos.), gårds-/gatenr (5 pos.) og bruks-/husnr (4 pos.). På landsbasis var det i 1998 og 1999 henholdsvis 1,5% og 1,2% av adressene som ikke koblet på 13 posisjoner. Som landsgjennomsnitt er dette forholdsvis bra, men det er store variasjoner mellom kommunene (SSB 99/1, SSB 1999/88).

Grunnen til at en valgte å koble på 13 posisjoner er at dette ga den beste koblingsprosenten. Ved å bruke et mer spesifisert adressebegrep, er det færre adresser som kobler. En optelling for 1.1.1998 viser at 4 % av adressene ikke kobler på 17 posisjoner (SSB 1999/88).

En kobling på 13 posisjoners adresse, dvs. uten å ta med feste-/oppgangsnummer, har imidlertid visse svakheter. Problemet gjelder i hovedsak matrikkeladresser, der eiendommer med samme bruksnummer kan dekke et forholdsvis stort område. Områder med samme bruksnummer er heller ikke alltid sammenhengende. Dersom det velges én koordinat for alle personene på samme gårds-/bruksnr kan resultatet bli svært misvisende.

For å få best mulig stedfesting foretas det **først** en kobling som inkluderer feste-/oppgangsnummer (17 posisjoner), for deretter å koble de resterende på 13 posisjoner. Dermed utnytter man koordinater der de er spesifisert på 17 posisjoner, og unngår at koblingen blir dårligere enn 13 posisjoners kobling.

Koblingen på 17 posisjoner for 1.1.2000 ga påført koordinat for ca. 4 050 000 personer (90,4%). For omtrent 307 000 av disse personene blir det en reell spesifisering på festenummer.

### 4.3. Koblingene videre

Personer som ikke får påført koordinater gjennom en kobling på 17 posisjoner blir behandlet videre med en kobling på 13 posisjoners adresse. Koblingen gjøres fremdeles mot adressedelen av GAB.

For dem som heller ikke får påført koordinat ved kobling på 13 posisjoner er det mulig å gjøre visse ekstrakoblinger. Dette går i korthet ut på å koble mot bygningsdelen og eiendomsdelen i GAB. Man prøver først å finne en koordinat til en bygning med samme adresse, og deretter prøver man å få kobling til en annen adresse på samme eiendom. Begge disse koblingene foregår på 13 posisjoners adresse.

Etter disse koblingene gjensto det for 1.1.2000-data fremdeles vel 112 000 personer uten koordinater. For om mulig å få koblet koordinat på disse tok vi utgangspunkt i adressen til de personene som bor på gateadresse (ikke matrikkel) og forsøkte å tilegne dem koordinaten til nabohusnummer. Et program sjekker husnummeret til begge sider for adressen (n-1 og n+1), og den nærmeste koordinaten som finnes tilegnes den aktuelle adressen. I dette steget brukes adresser med koordinater på inntil 5 nummer opp eller ned i forhold til utgangspunktet. Denne koblingen gav for 1.1.2000 ytterligere 0,6% av befolkningen koordinater. Hele koblingsprosessen er skissert i figur 4.1, og tallene for koblingene for data pr. 1.1.2000 vises i tabell 4.2.

**Tabell 4.2. Oversikt over koblingen i de ulike leddene i prosessen. Data fra 1.1.2000**

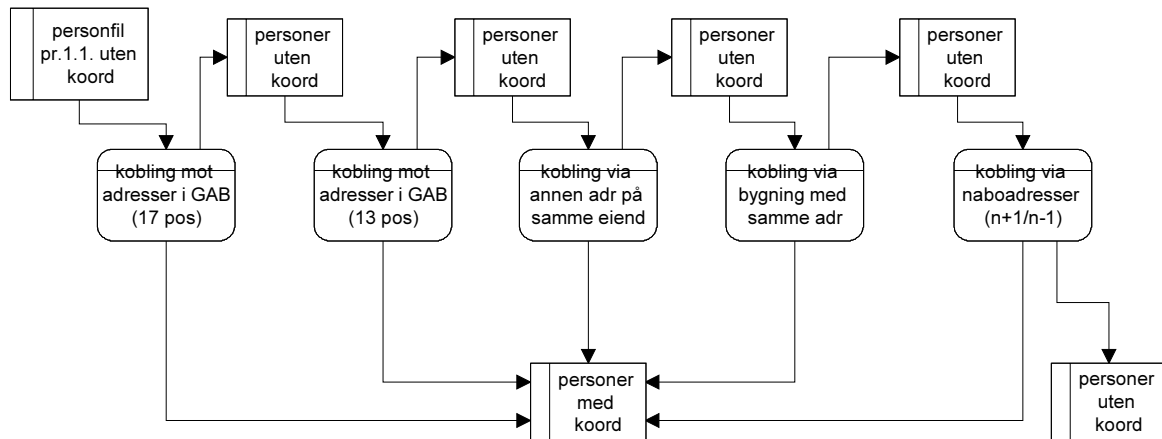
Kobling	Kode	Antall personer	Prosent av total
Totalt		4 478 473	100,0
17 pos. mot adressefil	S og K	4 047 977	90,4
- Rest		430 496	9,6
13 pos. mot adressefil	A	275 552	6,2
- Rest		154 944	3,5
13 pos. mot bygningsfil	B	42 897	0,9
- Rest		112 047	2,5
13 pos. mot eiendomsfil	E	0	0,0
- Rest		112 047	2,5
koord. til nabo 13 pos. gateadr.	N	27 598	0,6
Rest ikke koblet		84 449	1,9

Opplysninger om koblingene blir dokumentert. Dette sier noe om kvaliteten på koordinatene. Koordinatene kan også i noen tilfeller kan være feil i utgangspunktet. Dette **kan** være lett å avdekke

ved inntegning på kart (f.eks. hvis koordinaten ligger langt utenfor Norges grenser), men vil i mange tilfeller være umulig å avdekke uten befarings.

Når koordinatene er tilegnet bebodde adresser må de tilbakeføres til individfilen, for den aktuelle årgangen.

**Figur 4.1. Oversikt over koblingsprosessen for koordinatfesting**



Kommunevis oversikt over koblingsresultatene pr. 1. januar 1999 og 2000 finnes på internett (<http://www.ssb.no/befteft/tab-2001-12-06-07.html>).

## 5. Tilrettelegging av data

### 5.1. Lage rute-ID på befolkningsdataene

Selve tilretteleggingen av statistikken foregår i to deler. Først legges en rute-ID på befolkningsfilen, deretter hentes tabeller ut for det aktuelle området. Begge disse delene er skissert i figur 5.1.

Når koordinatfestingen av befolkningen er gjennomført, kan man starte tilretteleggingen av en rute-ID som skal knytte hvert individ på befolkningsfilen til en rute på kartet. Rute-ID lages med utgangspunkt i adressekoordinatene. Denne variabelen vil være forskjellig utfra hvor stor ruten er. Variabelen kalles RUxM, der x er antall meter i høyde og bredde på ruten.

Hovedprinsippet er at øst/vest-koordinaten og nord/sør-koordinaten rundes ned til nærmeste 1000 (når det er 1000 m-ruter) og slås sammen. Dette utgjør det nye identitetsnummeret. Prosessen programmeres på denne måten:

```
SET rute1000;
  oest1=((INT(oest_koo/1000))*1000);
  nord1=((INT(nord_koo/1000))*1000);
  LENGTH oest $ 9;
  LENGTH nord $ 9;
  oest=oest1;
  nord=nord1;
  oest2=RIGHT(oest);
  nord2=LEFT(nord);
  ru1000m=COMPRESS(oest2||nord2);
```

Programmet i sin helhet gjengis i vedlegg B.

Vi har nå laget en ny variabel på individfilen, som heter "ru1000m" og som er en ID for 1000-metersruter. Andre rutestørrelser som 100 m eller 250 m kan lages ved å bytte ut 1000-tallet.

### 5.2. Tilrettelegging av befolkningstabeller for ArcView

Statistikken og rutenettet kobles sammen i programvaren ArcView. Statistikktabellen må inneholde befolkningsopplysninger aggregert på ruter. Hver rute (rute-ID) skal ha en record som sier hvor mange som bor i ruten. Variablene i tabellen er:

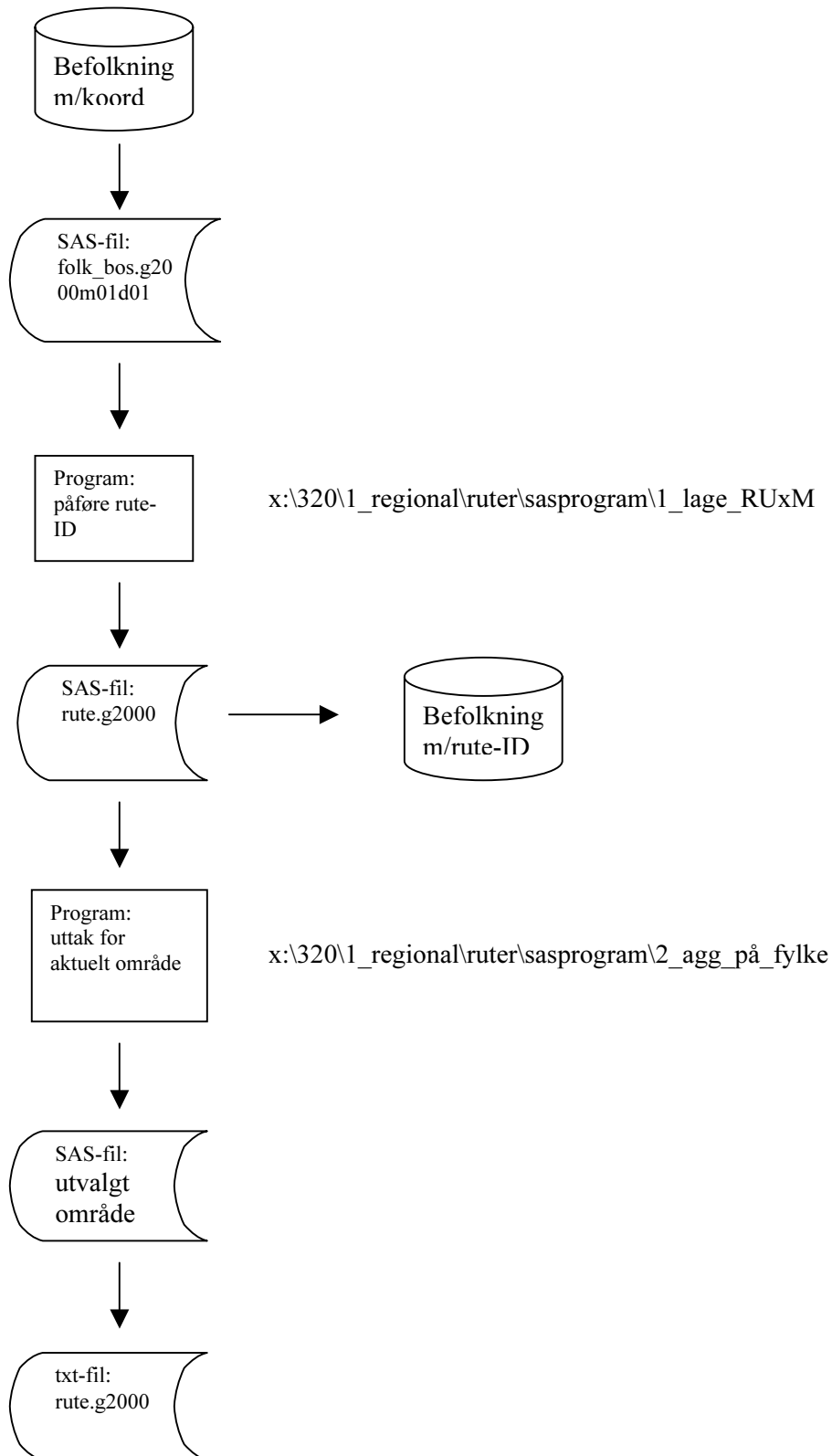
rute-id	total	menn	kvinner	0-5 år	6-15 år	16-19 år	20-39 år	40-66 år	67 +
---------	-------	------	---------	--------	---------	----------	----------	----------	------

Det kan være aktuelt å velge ut spesielle geografiske områder, f.eks. et fylke. Selv om rutene ikke følger fylkesgrensene er det mulig å velge ut ruter som inngår i et fylke. Ruter som ligger på grensen til et annet fylke bør da representeres med alle innbyggerne i ruten, og ikke kun de som bor i det ene fylket. Andre uttak, som ikke følger administrative inndelinger, kan gjøres direkte fra ArcView. Der har man muligheten til å velge akkurat de rutene man ønsker, gjerne utfra egne grensedata ("select by theme").

ArcView kan importere filer i text-format og i dBase-format. I dette prosjektet tilrettelegges tabellene i text-format, der linje 1 er variabelnavnene og der variablene er tabulator delt.

Programmet som bearbeider befolkningsfilen for kobling mot ArcView finnes i vedlegg B.

**Figur 5.1. Modellen skisserer arbeidet med å lage rute-ID på befolkningsfilen og å tilrettelegge tabeller for GIS**

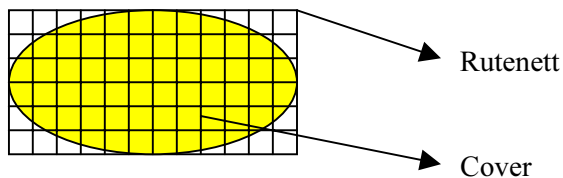


### 5.3. Lage rutenettet

Ettersom det er de samme rutenettene som brukes for ulike årganger av statistikken, må denne tilretteleggingen gjøres kun en gang. I forbindelse med dette prosjektet er det laget rutenett for hele landet på 1000 m og for byområdene Oslo, Bergen, Trondheim og Stavanger på 100 m. For andre størrelser eller områder må det lages et nytt rutenett. Dette arbeidet ligger ikke innenfor prosjektets primære rammer, men må vurderes i en utviklingsfase.

Selve rutenettet kan konstrueres som vektordata med programvaren ArcInfo. Et eget program ("rutenett.aml", vedlegg C) er blitt brukt til dette. I programmet angis størrelse på rutene, og ekstremverdiene for rutenettet angis ved henvisning til et cover slik at man får et rektangel med samme ekstremverdiene som coveret (se figur 5.2). Coveret som ble brukt til 1km-ruter for hele landet var i målestokk 1:2 000 000 (N2000).

**Figur 5.2. Rutenettet lages ut ifra ekstremverdiene i inncoveret**



Kommando for å kjøre AML-en:

```
&R rutenett [inncover som viser selve utstrekningen på nettet] [navn på utcover] <bredde> <høyde>
```

Parametrene for bredde og høyde settes i forhold til hvor store rutene skal være. For ruter på 1000x1000 meter skriver man altså:

```
&R rutenett [inncover som viser selve utstrekningen på nettet] [navn på utcover] 1000 1000
```

Deretter må man bygge topologi med kommandoen:

```
build [utcover] poly
```

Rutenettet kan nå hentes inn i ArcView. Ettersom rutenettet får en rektangulær utskrekning etter ytterpunktene vil det for 1000m-rutenettet for hele Norge bli enorme datamengder. Derfor er det hensiktsmessig å så tidlig som mulig i prosessen klippe vekk ruter som ikke dekker Norge. På grunn av at kartgrunnet (N2000) ikke har like høy detaljeringsgrad som bosettingskoordinatene buffres de digitale grensene for hele landet med en viss avstand. Deretter klippes rutenettet mot dette.

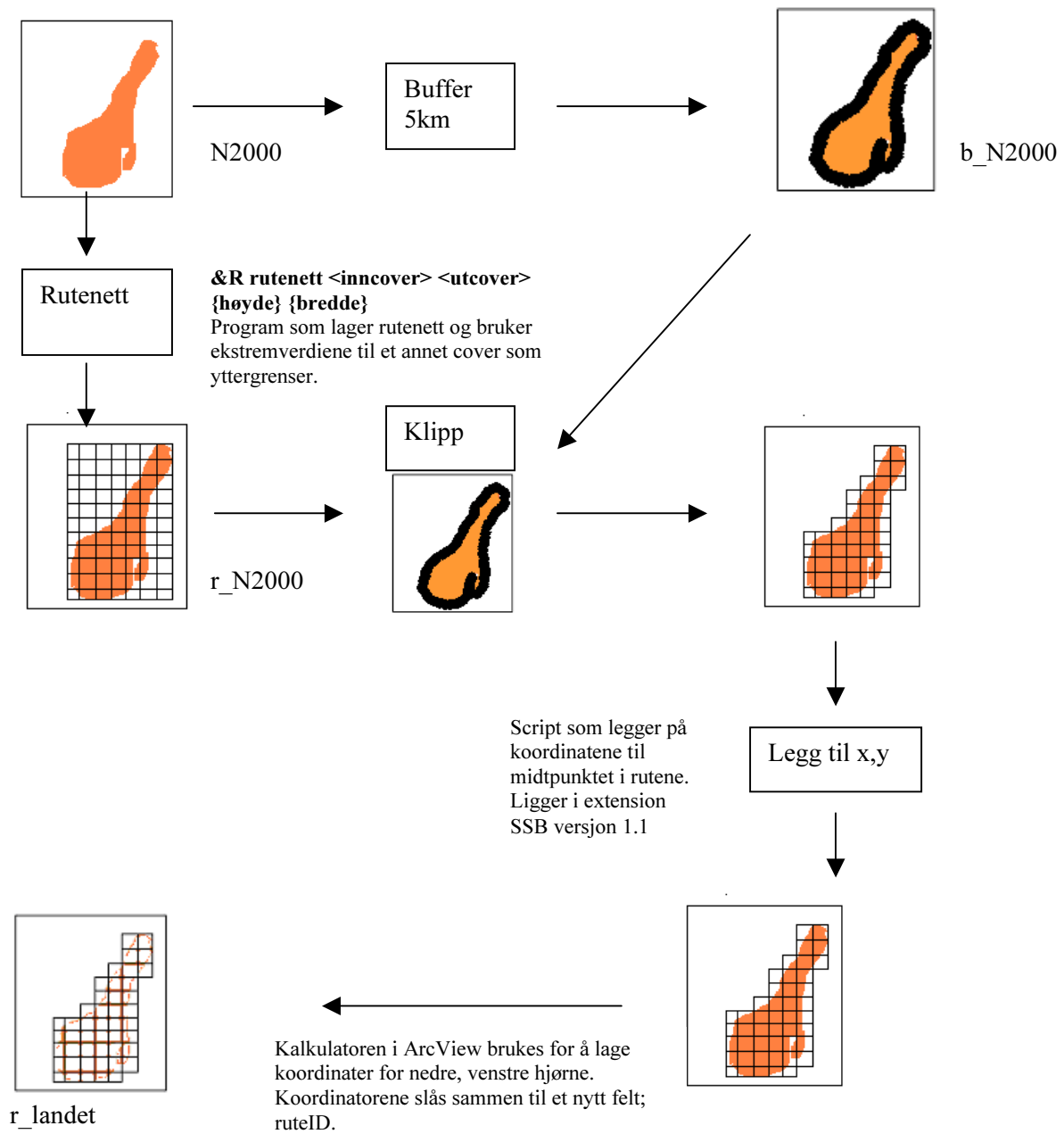
Koordinater tillegges rutene ved å hente inn en tilleggsfunksjon i ArcView (extension) som heter "SSB". Man får da opp funksjonen "legg til x,y". X-og Y-koordinaten som legges til i attributtabelen representerer et midtpunkt i rutene. Dette omgjøres til koordinatene i nedre, venstre hjørne ettersom det er det mest praktiske for den senere prosessen. Det er også nedre, venstre hjørne som brukes i statistikkbyråene i Sverige og Finland.

Det må da lages nye felt i tabellen der vi trekker fra 500 eller 50 meter (avhengig av hvor store rutene er). Disse nye verdiene defineres som tekststrenger. Deretter lages et nytt felt (ID), der øst/vest-koordinaten og nord/sør-koordinaten slås sammen;  $ID = x\_koor\ AsString + y\_koor\ AsString$

Det er dermed tilrettelagt for å koble statistikk til rutene via ID-en. Hele prosessen er skissert i figur 5.3.



**Figur 5.3. Skisse som viser hvordan selve rutenettet blir laget. Det samme prinsippet gjelder for rutenett av andre størrelser og for andre områder**



## 5.4. Produksjonslinje

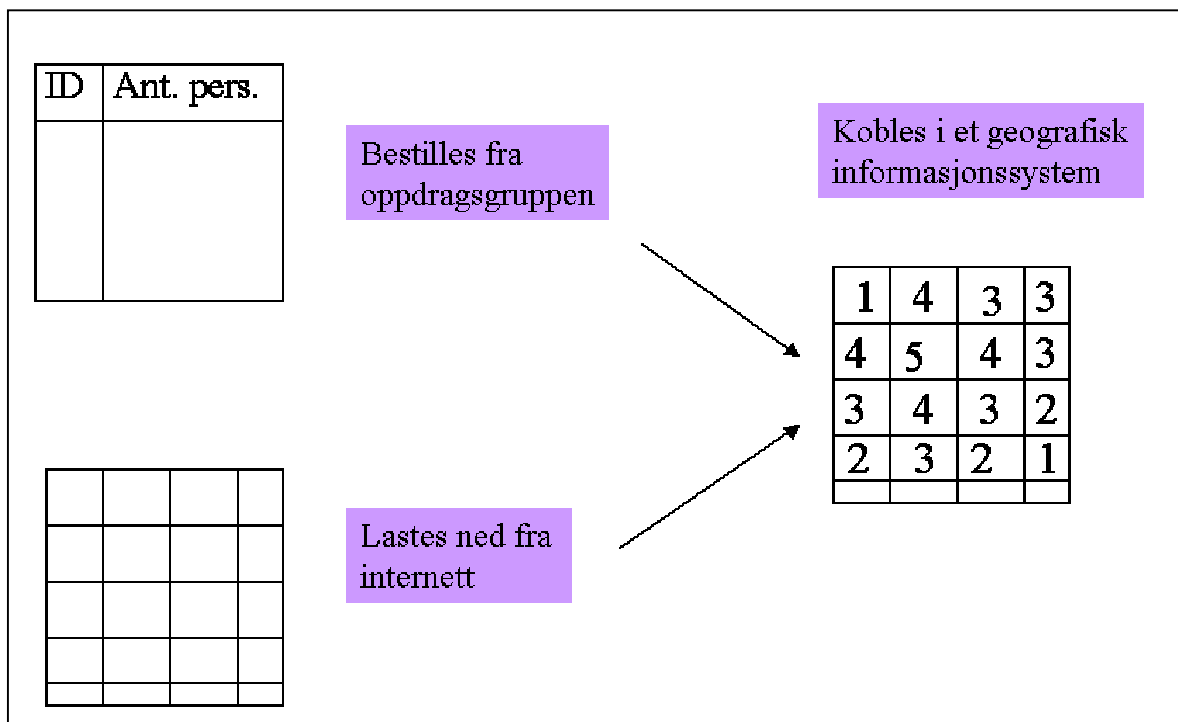
Ved å velge alternativ 1 (kapittel 3.2.) kan rutenettet og statistikken ses på som to separate deler. Gjennom koblingsnøkkelen, en felles ID på rutenettet og i statistikken, kan de to delene knyttes til hverandre i et GIS.

Når brukerne har et rutenett som dekker det aktuelle området, kan de gjøre koblingen til statistikk-tabellene selv. I prinsippet vil kundene kunne lage et rutenett selv, men SSB bør legge opp til et mer brukervennlig opplegg. Det vil være en fordel om brukerne får tilgang til rutenett med rute-ID direkte fra SSB. På grunn av de store datamengdene vil den enkleste måten å løse dette på, være å legge et nedlastbart rutenett ut på internett.

Ettersom et slikt rutenett ikke vil inneholde noe statistikk og lages som en engangsjobb, er det rimelig at det er gratis. Dette vil også være den løsningen som er minst tidkrevende for SSB. Det vil da kun være selve statistikken som SSB må tilrettelegge for kundene. Statistikktabellene bestilles fra Seksjon for befolknings- og utdanningsstatistikk sin oppdragsgruppe. I kapittel 6 diskuteres hvilke hensyn som må tas til personvernet i forbindelse med rutenettstatistikk.

Den foreslåtte produksjonslinjen skisseres i figur 5.4.

Figur 5.4. Produksjonslinjen for rutenettstatistikk



## 6. Hensynet til personvern

### 6.1. Problemstilling

Løpende produksjon av statistikk knyttet til rutenett er nytt i SSB. Med en rutestørrelse på 1000 x 1000 m, vil rutestatistikken gi atskillig mer detaljerte opplysninger enn de fleste grunnkretser. Det gjelder særlig i spredtbygde områder, der grunnkretsene som regel er av stor utstrekning. Det er derfor naturlig å trekke spørsmålet om personvern inn i dette prosjektet.

*Alder og kjønn* er de personopplysningene som i første omgang vil inngå i rutenettproduktet. Derfor er det først og fremst disse variablene som vil bli diskutert her.

Som hovedregelen skal SSB ikke offentliggjøre opplysninger om enkeltpersoner. Til forskning og offentlig planlegging formidles det enkelte ganger individdata, men da i all hovedsak i helt anonymisert form. Unntaksvis kan det være aktuelt med formidling av identifiserbare individdata til forskning, men da skal forskeren og SSB ha fått godkjent dette av Datatilsynet. Formidling av statistikk fra SSB kan altså deles i to hovedtyper; offentliggjøring og utlevering. Når statistikk offentliggjøres vil det si at data gjøres tilgjengelig for alle, mens ved utlevering er det kun visse brukere som får tilgang til dataene.

### 6.2. Utlevering av opplysninger

Adgangen til å *utlevere* opplysninger fra Statistisk sentralbyrås statistiske registre følger av statistikkloven § 2-5 og Statistisk sentralbyrås rammekonsesjon fra Datatilsynet pkt 2.4. I Sikkerhetshåndbok for Statistisk sentralbyrå per juni 2001 heter det bl.a. at:

- Utlevering av individualopplysninger kan skje til etablerte forskningsinstitutter, forskere med forskningskompetanse og hovedfagsstudenter som er under veiledning av en med forskerkompetanse ved undervisningsstedet.
- Utlevering til offentlig planlegging kan bare skje til organer i stat eller kommune.
- Utlevering betinger bruk i konkret prosjekt.

Ved utlevering kan personopplysningene både være identifiserbare og aidentifiserte. I forbindelse med rutenettdata er det kun snakk om aidentifiserte data, altså data som *indirekte* kan føres tilbake til enkeltpersoner. Når punktene over er oppfylt kan altså brukerne søke Datatilsynet og SSB om å få utlevert opplysninger knyttet til rutenett.

Mottaker av opplysningene må forholde seg til personopplysningslovens regler, dvs. søke konsesjon dersom det er tale om sensitive personopplysninger (jf pol § 33) og melde behandlingen i forkant dersom det er tale om ikke-sensitive personopplysninger (jf pol § 31)

### 6.3. Offentliggjøring av opplysninger

Når det gjelder offentliggjøring av opplysninger er hovedregelen at opplysninger ikke i noe tilfelle skal offentliggjøres slik at de kan føres tilbake til oppgavegiver eller annen identifiserbar enkeltperson. Dersom det unntaksvis - av hensyn til hensiktsmessig oppbygging av statistikken - er nødvendig at enkelte oppgavegivere eller andre likevel blir identifisert, er det absolutt krav at slik identifisering ikke er til skade for fysisk person eller til urimelig skade for juridisk person.

Før rutenettproduktet kan offentliggjøres må altså dataene behandles med tanke på å unngå at opplysninger indirekte skal kunne tilbakeføres til en identifiserbar enkeltperson. Spørsmålet blir så: Når kan man si at opplysninger kan tilbakeføres til identifiserbare enkeltpersoner?

SSB har som hovedregel at det ikke skal publiseres tabeller med mindre enn tre enheter i en tabellcelle hvor utvalget medfører fare for identifisering. Faren for at opplysninger skal kunne knyttes til ett individ regnes som forholdsvis stor dersom verdien er 3 eller færre. Datatilsynet legger også til grunn i saker hvor personverntilsynet er fremtredende, at opplysninger som kan føres tilbake til en enhet på mindre enn fire personer ikke er anonym. Tilsvarende praksis har Rikstrykdeverket som legger til grunn at persontall under 4 medfører fare for personidentifisering.

Under blir det foretatt en vurdering av hvilke metoder som kan brukes i rutenettprosjektet for å unngå identifisering i tabellceller med få enheter.

## 6.4. Aldersgruppering

En mye brukt metode for å redusere sjansene til å ende opp med få personer i en celle, er å slå sammen verdier i grove inndelinger, såkalt global omkoding (SSB 2001/28). Global omkoding av aldersvariabelen i rutenettprosjektet er hensiktsmessig ikke bare pga. konfidensialitetshensyn, men også pga. store datamengder. Det vil da være naturlig å benytte aldersgrupperinger som skiller ut barn under skolealder, skolebarn, ungdom, voksne og eldre.

En standard for aldersgruppering brukt i SSB er: 0-5 år, 6-15 år, 16-19 år, 20-66 år og 67 år og mer. Av standard aldersgrupperinger brukt i SSB er denne den mest hensiktsmessige for dette produktet. Det foreslås imidlertid at den store gruppen mellom 20 og 66 år deles i to: fra og med 20 til og med 39 og fra og med 40 til og med 66. På sikt kan også andre aldersgrupper vurderes. Det kan f.eks. også være et poeng å skille mellom elever i barneskolen og ungdomsskolen.

På tross av den ovenfornevnte omkodingen vil datamaterialet fremdeles inneholde tabellceller med lav verdier. Tabell 6.1 illustrerer situasjonen for Kongsvinger kommune ved en rutestørrelse på 1000 m x 1000 m. Mens 10 % av rutene har kun én innbygger totalt sett, blir andelen tabellceller med verdien 1 44 % når variablene kjønn og aldersgrupper er med.

**Tabell 6.1. Antall celler med cellefrekvens (ant. personer) 1 til 8, når Kongsvinger kommune er fordelt på ruter, ruter x kjønn, ruter x alder og ruter x kjønn x alder. Prosentandelene er regnet av antall celler med cellefrekvens minst 1. Tallene i fet kursiv i kolonnen lengst til høyre er summen av tallene til venstre for det**

Originale frekvenser (verdien i cellen)	Cellefrekvenser							
	1	2	3	4	5	6	7	8
Antall og andel ruter m bef.	40 10,2%	34 8,7%	25 6,4%	20 5,1%	23 5,9%	26 6,6%	17 4,3%	207 52,8%
Antall personer	40	68	75	80	115	156	119	<b>653</b>
Rute x kjønn	138	95	68	59	59	43	24	252
Antall og andel celler	18,7%	12,9%	9,2%	8,0%	8,0%	5,8%	3,3%	34,1%
Antall personer	138	190	204	236	295	301	168	<b>1532</b>
Rute x alder	397	316	137	123	77	66	39	260
Antall og andel celler	28,1%	22,3%	9,7%	8,7%	5,4%	4,7%	2,8%	18,3%
Antall personer	397	632	411	492	385	396	273	<b>2986</b>
Rute x alder x kjønn	1010	454	244	119	87	53	29	302
Antall og andel celler	44,0%	19,8%	10,6%	5,2%	3,8%	2,3%	1,3%	13,1%
Antall personer	1010	908	732	476	435	318	203	<b>4082</b>

## 6.5. Prikking

Prikking eller "cell suppression" er ved siden av global omkoding en av de vanligste måter å hindre at sensitive opplysninger skal kunne avsløres gjennom tabeller. Dette innebærer at lave verdier i tabellceller undertrykkes. Når prikking brukes i tabeller som inneholder marginaler, ønsker en ofte å publisere marginalene dersom disse i seg selv ikke oppfattes som konfidensielle. For at innholdet av den prikkede cellen ikke skal kunne utledes av marginalene og de øvrige celletallene, må også andre celler prikkes. Dette kalles sekundærprikking (SSB, Notater 2001/28).

I rutenettprosjektet kan det tenkes at alle celler med under 4 i verdi prikkes. Tabell 6.1 viser hvordan antall celler med få observasjoner øker når antall variable øker. Den viser også hvordan antall celler synker med høyere celfrekvens, en tendens som fortsetter også for celler med høyere frekvenser enn dem som vises i tabell 6.1. Dette innebærer at prikking av tabellceller med lave verdier vil føre til at svært mange av verdiene undertrykkes.

Et av hovedpoengene med rutenettprosjektet er at man skal kunne benytte rutene som byggeklosser for egne aggregeringer. Når verdier i aggregatene er prikket, vil dette føre til avvik fra det opprinnelige sumtallet. Jo større aggregeringene er, desto større vil avviket fra det reelle innbyggertallet være. Tabell 6.2 er et utvalg av ruter fra Kongsvinger kommune. Befolkningen er kun fordelt på aldersgrupper. Ettersom svært mange av tabellcellene inneholder verdier som er lavere enn 4, mister vi mye informasjon. I aldersgruppen 0-5 år bor det totalt 9 personer, men ettersom det kun er 3 og færre i hver tabellcelle blir resultatet etter prikking 0.

Et alternativ til å prikke alle tabellcellene kunne vært å beholde totalsommene uprikket. Brukerne vil allikevel miste så mye informasjon om alders- og kjønnsfordeling at nytteverdien av produktet kan diskuteres. Prikkemethoden vurderes som uegnet for bruk i rutenett-prosjektet.

**Tabell 6.2. Befolkningsfordeling på et utvalg av ruter i Kongsvinger kommuner. Før og etter prikking**

rute_id	Total		0-5		6-15 år		16-19 år		20-39 år		40-66 år		67 år +	
	original	prikket	original	prikket	original	prikket	original	prikket	original	prikket	original	prikket	original	prikket
3270006676000	57	57	2	..	8	8	0	..	10	10	15	15	22	22
3270006677000	21	21	2	..	1	..	1	..	3	..	5	5	9	9
3270006694000	2	..	0	..	0	..	0	..	0	..	1	..	1	..
3280006665000	8	8	0	..	0	..	1	..	4	4	2	..	1	..
3280006674000	1	..	0	..	0	..	0	..	0	..	0	..	1	..
3280006675000	2	..	0	..	0	..	0	..	0	..	1	..	1	..
3280006676000	25	25	3	..	6	6	0	..	4	4	11	11	1	..
3280006677000	46	46	2	..	4	4	2	..	17	17	16	16	5	5
3280006678000	1	..	0	..	0	..	0	..	0	..	0	..	1	..
3280006680000	2	..	0	..	0	..	0	..	0	..	2	..	0	0
<b>10 ruter</b>	<b>165</b>	<b>157</b>	<b>9</b>	<b>0</b>	<b>19</b>	<b>18</b>	<b>4</b>	<b>0</b>	<b>38</b>	<b>35</b>	<b>53</b>	<b>47</b>	<b>42</b>	<b>36</b>

**Tabell 6.3. Befolkningsfordeling på et utvalg av ruter i Kongsvinger kommuner. Før og etter kontrollert ikke-stokastisk avrunding (Argus)**

rute_id	Total		0-5		6-15 år		16-19 år		20-39 år		40-66 år		67 år +	
	original	avr	original	avr	original	avr	original	avr	original	avr	original	avr	original	avr
3270006676000	57	55	2	5	8	5	0	0	10	10	15	15	22	20
3270006677000	21	25	2	5	1	5	1	0	3	0	5	5	9	10
3270006694000	2	5	0	0	0	0	0	0	0	0	1	5	1	0
3280006665000	8	10	0	0	0	0	1	0	4	5	2	0	1	5
3280006674000	1	5	0	0	0	0	0	0	0	0	0	0	1	5
3280006675000	2	0	0	0	0	0	0	0	0	0	1	0	1	0
3280006676000	25	25	3	5	6	5	0	0	4	5	11	10	1	0
3280006677000	46	45	2	0	4	5	2	0	17	20	16	15	5	5
3280006678000	1	0	0	0	0	0	0	0	0	0	0	0	1	0
3280006680000	2	5	0	0	0	0	0	0	0	0	2	5	0	0
<b>10 ruter</b>	<b>165</b>	<b>175</b>	<b>23</b>	<b>15</b>	<b>19</b>	<b>20</b>	<b>4</b>	<b>0</b>	<b>38</b>	<b>40</b>	<b>53</b>	<b>55</b>	<b>42</b>	<b>45</b>

## 6.6. Avrunding

Avrunding vil si å erstatte en celledverdi med en avrundet verdi utifra et visst sett med kriterier. Avrunding av celfrekvens  $x$  vil si å erstatte  $x$  med et tall  $kb$  der  $b$  er en på forhånd valgt base, f.eks. 3, 5 eller 10 og  $k$  søkes valgt slik at  $kb$  ligger så nær  $x$  som mulig. Vanligvis vil  $k$  være  $\lceil x/b \rceil$ , det største heltall mindre enn  $x/b$  (avrunding nedad) eller  $\lfloor x/b \rfloor + 1$  (avrunding oppad).

Det er mange hensyn som må tas i forbindelse med avrunding, skal f.eks. verdiene rundes opp eller ned, hva skal basen være, skal både høye og lave verdier avrundes, hvordan skal marginalene behandles. Disse temaene diskuteres grundig av Johan Heldal i foreløpig upublisert [notat om avrunding i frekvenstabeller](#). Tabell 6.3 er et eksempel på avrunding der basen er 5, marginalen ikke avrundet mer enn basen, samtidig som de indre cellene i tabellen adderer seg opp til marginalene.

Hovedpoenget for rutenett-prosjektet må være at avviket mellom avrundede og urundede tall skal være så lite som mulig, uten at opplysninger kan tilbakeføres til enkeltpersoner. Det gjelder både enkeltceller og marginalene. En sammenligning av tabell 6.2 og tabell 6.3 viser at avrundingen gir et resultat som er mer i overensstemmelse med de opprinnelige tallene.

Den avrundingsmetoden som regnes som den beste er kontrollert ikke-stokastisk avrunding (brukt i tabell 6.3). Programvaren Argus, som er utviklet av statistikkbyrået i Nederland, foretar avrunding etter denne metoden. Til tross for gode resultater, viste det seg at datamengdene i rutenettprosjektet var for store for Argus.

Det nestbeste alternativet for avrunding av rutenettdata er kontrollert stokastisk avrunding. Pt. arbeides det med å lage et program som kan foreta denne avrundingen. Programmet lages i SAS, og det vil derfor kjøres i samme hånd vending som SAS-programmet for tilrettelegging av rutenett-tabellene.

## 6.7. Flere variabler?

I tillegg til alder og kjønn, kan også andre variabler kobles til produktet. Av variabler tilknyttet Seksjon for befolknings- og utdanningsstatistikk kan nevnes sivilstatus, statsborgerskap, landbakgrunn og utdanningsnivå. (Produktet kan i teorien utvides med alle variable som kan knyttes til

fødselsnummer.) De nevnte variablene er av en annen karakter enn alder og kjønn, selv om heller ikke disse er sensitive etter personopplysningsloven. Opplysningene kan karakteriseres som *beskyttelsesverdige opplysninger* (SSB, Notater 2001/28) ettersom en del mennesker vil ønske at de ikke skal kunne leses av alle.

Det foreslås imidlertid at prosjektet i første omgang begrenses til variablene alder og kjønn, slik at erfaring med bruk av denne type datasett kan gjøre oss bedre i stand til å foreta de riktige valgene for andre variable senere. Dersom det blir vanskelig å ivareta konfidensialitetshensynet kan én mulig løsning bli at et utvidet produkt må behandles som utlevering av data, med egne avtaler. Men dersom SSB tilbyr brukerne rutenett med antall personer totalt, kjønnsfordeling og aldersgruppering vil de mest generelle behovene være tilfredsstillt.

## 6.8. Størrelse på rutene

Når det gjelder størrelsen på rutene vil mindre ruter øke mulighetene for identifisering ettersom det blir færre personer innenfor samme område/rute. Dette er imidlertid avhengig av hvor mange som bor i hver rute, altså hvor tett folk bor. Det foreslås her at det lages et rutenett som gjelder hele landet med rutestørrelse på 1 x 1 km. Dette er en rutestørrelse som er godt egnet for kart for større områder, som f.eks. hele Norge. Det kan også gi et greit overblikk i over bosettingsmønsteret i mindre områder. 1 km<sup>2</sup>-ruter brukes også i nordisk felles prosjekt for rutenettkart.

Til sammenligning med grunnkretsene kan det nevnes at nærmere 9 000 av de rundt 14 000 grunnkretsene har en utstrekning på over 1 km<sup>2</sup>. Ca. 3 300 grunnkretser er større enn 20 km<sup>2</sup>.

For de tettest befolkede områdene vil det i mange sammenhenger være ønskelig med mer detaljerte opplysninger, f.eks. 100 m x 100 m rutenett. Det er også i disse områdene utstrekningen på grunnkretsene er minst. De tettest befolkede områdene i Norge samsvarer til en viss grad med de kommunene med høyest folke mengde. Dersom det skal frigis mer detaljerte data enn 1 km<sup>2</sup>, er det naturlig i første omgang å starte med detaljerte rutenett for Oslo-, Bergen-, Stavanger- og Trondheimsområdet.

På sikt bør det også vurderes om mer detaljerte rutenettdata også skal tilbys for andre områder. Hvor er det naturlig at grensen går? I denne sammenhengen bør det også tas hensyn til hvor ressurskrevende tilretteleggingen blir. Et rutenett på 100m<sup>2</sup> for hele landet vil gi en fil på enorme datamengder, og det vil derfor ikke være hensiktsmessig å legge en slik fil ut på internett for nedlasting. Derfor bør 100m x 100m rutenettet som legges ut på internett kun gjelde for de enkelte områdene. For ikke å måtte legge ut veldig mange filer, kan det tenkes at spesielt interesserte kan få muligheten til å få *statistikken* på 100m x 100m-ruter, men at de må konstruere rutenettet selv. På sikt kan det vurderes å utvide det geografiske området for 100m-ruter, ev. å innføre en annen rutestørrelse, f.eks. 250 m.

## 6.9. Oppsummering

Rutenett-prosjektet innbefatter i første omgang variablene alder og kjønn. Aldersvariabelen grupperes for å redusere sjansene for tabellceller med få personer. Dataene bør viderebehandles for å unngå at opplysninger kan tilbakeføres til identifiserbare personer. Den metoden som er best egnet i forbindelse med rutenett-statistikk er avrunding. Et program for avrunding er under utvikling.

I spesielle tilfeller kan brukere få tilgang til rutenett-data som ikke er avrundet. En slik tillatelse gis etter §2-5 i Statistikkloven.

Størrelsen på rutene er 1 km x 1 km for hele landet, men 100 m x 100 m for de 4 største byområdene. På sikt vurderes mulighetene for å gi 250 m x 250 m ruter for andre områder.

## 7. Noen resultater

### 7.1. Ferdig produkt og testbruk

Innenfor dette prosjektet tilrettelegges det for visse standardprodukter. Produktene vil bestå av rutenett, som brukerne selv kan laste ned fra internett, og statistikk som bestilles fra SSB. Statistikk for 1 km x 1 km vil foreligge for hele landet, og brukerne kan bestille for hele landet eller fylkesvis. Statistikk for 100 m x 100 m ruter vil foreligge for storbyområdene, Oslo, Bergen, Trondheim og Stavanger.

Variablene som knyttes til er kjønn og alder. Alder er gruppert og aldersgruppene som er brukt er 0-5 år, 6-19 år, 20-39 år, 40-66 år og 67 år og over.

For å vurdere produktets nytte og brukervennlighet ble det trukket inn en testbruker av dataene. Statens vegvesen Akershus (SvA) fikk i oktober 2001 tilgang til 100m-rutenett for Oslo-området med statistikk. Dette ville de bruke i forbindelse med planlegging av infrastruktur for kollektivtrafikken. I dette inngikk blant annet vurdering av stoppesteder, terminaler og innfartsparkeringsanlegg for sykkel og privatbil. Planleggingen var særlig knyttet til utredning av Kolsåsbanen. Dataene var ikke behandlet med tanke på konfidensialitet, men ble utlevert i henhold til §2-5 i Statistikkloven.

SSB ønsket en tilbakemelding fra SvA på følgende punkter;

- Ev. tekniske problemer i forbindelse med bruk av produktet
- Utvalg av område i forhold til at det skal dekke "Oslo-området"
- Nytteverdi
- Eksempler på bruk, kort omtale og noen illustrasjoner
- Eventuelt andre ting

Det ble holdt et evalueringsmøte i desember 2001 der disse temaene ble diskutert.

Den generelle tilbakemeldingen fra SvA var at de var svært fornøyd og hadde hatt god nytte av dataene. Den tekniske delen som må på plass før dataene kan analyseres består av å koble statistikken til rutenettet. Dette hadde vært uproblematisk, selv om brukeren ikke hadde mye erfaring med programvaren ArcView.

100 m x 100 m-ruter for Osloområdet gjelder kun de kommunene som Oslo tettsted inngår i. SvA syntes i utgangspunktet dette var greit, men i forbindelse med andre vegprosjekter kunne de også tenkt seg rutenett-data over andre områder, f.eks. området rundt Gardermoen flyplass. De hadde imidlertid forståelse for at de mest detaljerte dataene kun gjelder i de tettest bebodde områdene. Et mulig fremtidig rutenett på 250 x 250 m kan være godt egnet for slike analyser i mindre tettbebygde områder. Km-ruter vil antageligvis ikke være detaljert nok i denne sammenheng.

SvA hadde benyttet dataene til å se på bl.a.:

- befolkningskonsentrasjoner i forhold til jernbanestasjoner
- behovet for innfartsparkering reflektert ved kart som viser befolkningen bosatt innenfor/utenfor en buffer på 400m i forhold til busslinje
- bosettingsmønsteret i forhold til foreslåtte busslinjer
- gjennomsnittlig befolkningsgrunnlag i luftlinjeavstand fra stoppestedene langs ulike baner

SvA hadde ikke benyttet kjønns- og aldersvariablene i produktene, men så ikke bort i fra at det kunne være aktuelt i noen sammenhenger. SvA var interessert i flere data fra SSB knyttet til rutenett. Blant annet ble det nevnt biler og arbeidsplasser.



## 7.2. Sammenligning med andre nordiske land

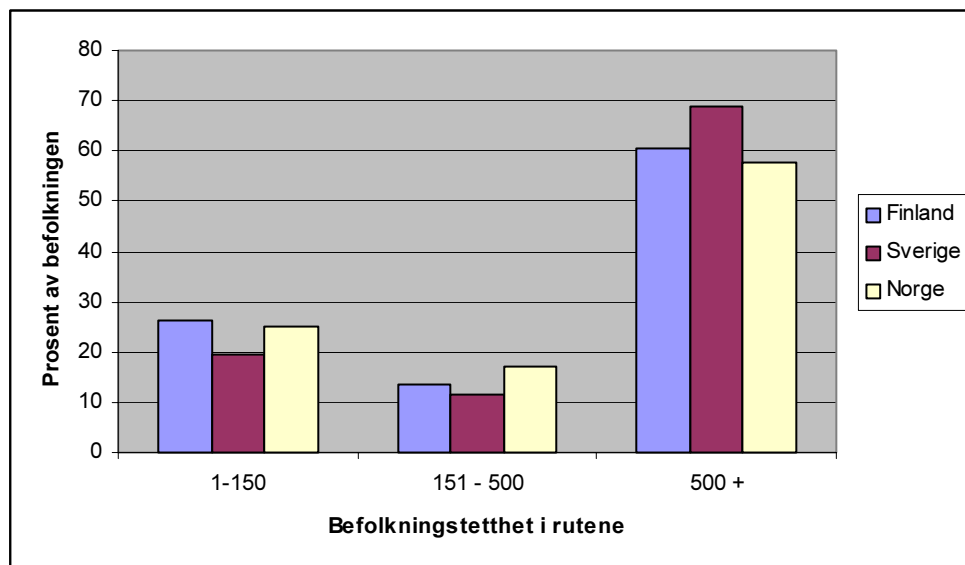
Ettersom rutenettstatistikk allerede er etablert i det svenske og det finske statistikkbyrået, kan det være interessant å sammenligne noen hovedresultater med disse landene. I 1995 ble det utarbeidet en rapport for et samarbeidsprosjekt mellom Sverige og Finland (Statistics Finland 1995). Sammenligningene tar utgangspunkt i tabeller fra denne rapporten, der tall for Sverige og Finland er fra 1995. Tallene for Norge gjelder per 1.1.2000.

**Tabell 7.1. Fordelingen av befolkningen og ruter i forhold til hvor tett bebodde rutene er. 3-delning**

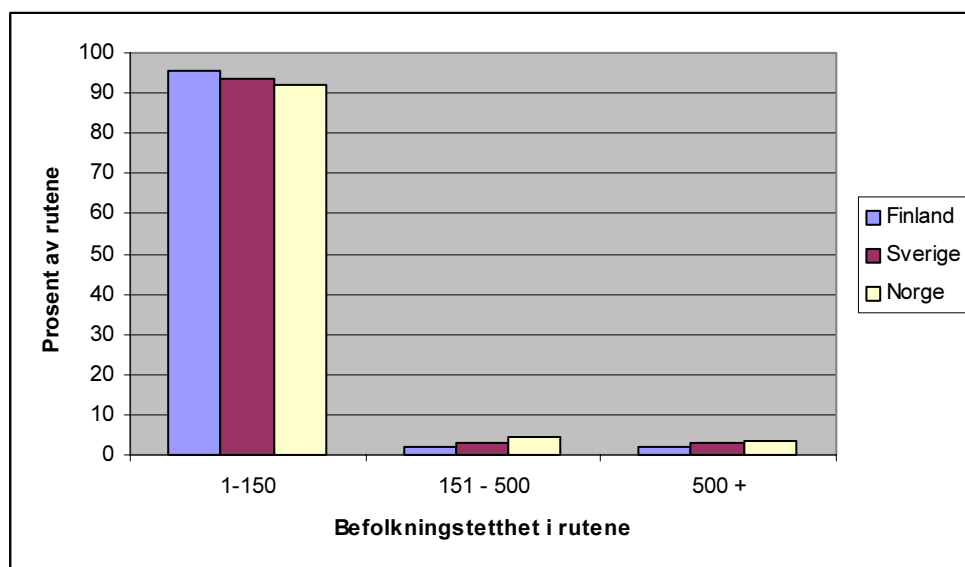
Befolkningstetthet (innb./km <sup>2</sup> )	% av befolkningen			% av rutene		
	Finland	Sverige	Norge	Finland	Sverige	Norge
1-150	26,1	19,4	25,1	95,7	93,7	92,1
151 - 500	13,6	11,6	17,0	2,2	3,1	4,6
500 +	60,3	69,0	57,9	2,0	3,2	3,3
Totalt	100,0	100,0	100,0	100,0	100,0	100,0

Tabell 7.1 viser at det i Sverige er en forholdsvis lav andel av befolkningen som bor i ruter med få personer (1-150). Finland og Norge har en andel som er rundt 6 prosentpoeng større. Norge har den laveste andelen av befolkningen som bor i tett bebodde ruter (mer enn 500 personer). Til gjengjeld er disse personene fordelt på den høyeste andelen ruter av de tre landene. Se også figur 7.1 og figur 7.2.

**Figur 7.1. Fordelingen av befolkningen i forhold til hvor tett bebodd rutene er. 3-delning**



**Figur 7.2. Fordelingen av rutene i forhold til hvor tett bebodd rutene er. 3-delning**



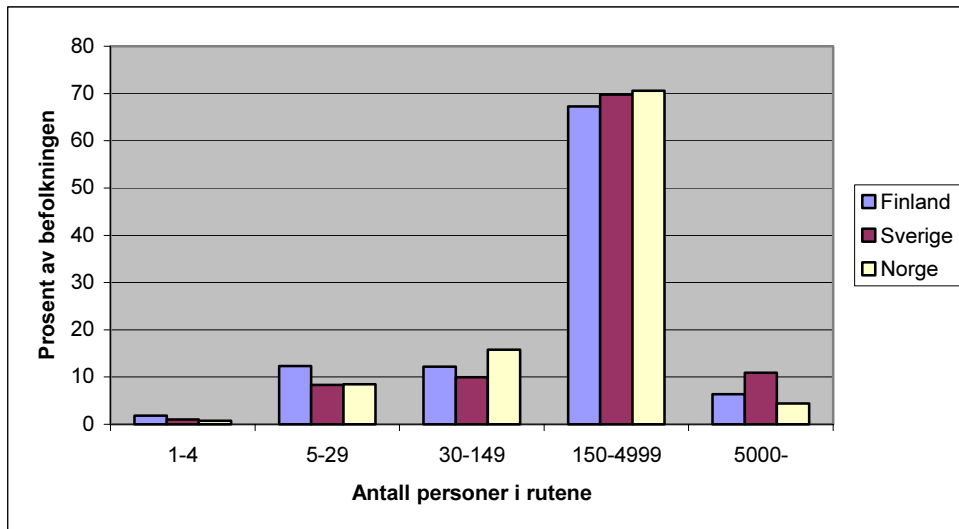
Alt i alt kan man etter denne grove sammenligningen mellom disse 3 nordiske landene imidlertid si at mønsteret er ganske likt.

Foretar man den samme sammenligningen, men med en finere inndeling av befolkningstettheten, blir forskjellene mer tydelige (se tabell 7.2). Av de tre landene har Norge lavest andel personer i ruter med 1-4 innbyggere og over 5000 innbyggere (figur 7.3). Sverige har høyest andel personer og ruter med over 5000 innbyggere. Arealmessig er det for alle tre landene ruter med mellom 5 og 29 innbyggere som dominerer kartet (figur 7.4).

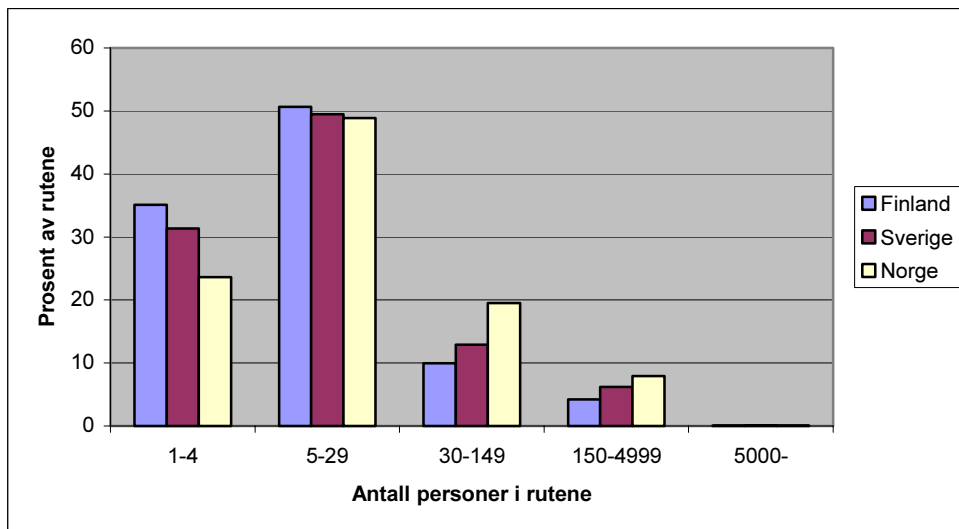
**Tabell 7.2. Fordelingen av befolkningen og ruter i forhold til hvor tett bebodde rutene er. 5-delning**

Befolkningstetthet (innb./km <sup>2</sup> )	% av befolkningen			% av rutene		
	Finland	Sverige	Norge	Finland	Sverige	Norge
1-4	1,83	1	0,76	35,1	31,35	23,64
5-29	12,35	8,33	8,50	50,67	49,47	48,89
30-149	12,16	9,99	15,77	9,97	12,88	19,50
150-4999	67,29	69,76	70,59	4,21	6,19	7,93
5000-	6,38	10,91	4,38	0,04	0,11	0,04

**Figur 7.3. Fordelingen av befolkningen i forhold til hvor tett bebodd rutene er. 5-delning**



**Figur 7.4. Fordelingen av rutene i forhold til hvor tett bebodd rutene er. 5-delning**



## 8. Oppsummering

### 8.1. Muligheter videre fremover

Ettersom vi har koordinat på hver person (98%) i Norge, vil vi teoretisk kunne lage rutenett-statistikk på all informasjon som kan kobles til fødselsnummer. Mulighetene er altså store; utdanning, sivilstatus, landbakgrunn, inntekt etc. En utvidelse av tilbudet vil imidlertid forde stor forsiktighet med hensyn til personvernet og muligheter for misbruk. Det bør også være påvist en relevans av at statistikken presenteres på denne formen.

Også på andre områder enn personstatistikk kan rutenett-produkt være relevant. SSB har nå avsluttet prosjektet med stedfesting av bedrifter og det foreligger derfor koordinat på ca. 75% av alle bedriftene i Norge. Dermed har man mulighet til å lage rutenett-statistikk på bedrifter. Et rutenett-kart over antall ansatte i bedriftene vil f.eks. kunne være veldig nyttig i mange sammenhenger, blant annet for å synliggjøre transportbehovet for ulike områder (jf. kap. 7.1 om testbruker).

Et mulig utviklingsprosjekt kan være å tilby en pakke med forskjellige typer rutenett-statistikk. Man kan også tenke seg en CD-ROM som også inneholder forskjellige typer kartgrunnlag (inkl. rutenett) og med en enkel utgave av et GIS-verktøy. En slik CD-ROM måtte antakelig vært utviklet i samarbeid med Statens kartverk (kartgrunnlag) og programvare-distributør. Det finske statistikkbyrået har blant annet laget et slikt produkt (TeemaCD), og SSB bør eventuelt dra nytte av erfaringer de har gjort seg.

## **Referanser**

Statistics Finland (1995): "Differences in the spatial structure of the population between Finland and Sweden in 1995". Research report 221.

Heldal, Johan og Johan Fosen (2001): "Statistisk konfidensialitet i SSB. Et diskusjonsnotat", Notater 2001/28, Statistisk sentralbyrå.

## Vedlegg A: Filstruktur

x:\320\1\_regional\ruter

<b>-g2000</b>	(statistikk på ruter fra 1.1.2000)
<b>-fNN</b>	(en katalog med fylkesfiler)
<b>-01_1000</b>	(filer for hvert fylke_meter)
<b>-01_250</b>	
<b>-02_1000</b>	
<b>-osv</b>	
<b>-landet</b>	(en fil for hele landet, 1000m)
<b>-oslo</b>	(en fil for Oslo, 100m)
<b>-bergen</b>	(en fil for Bergen, 100m)
<b>-stavanger</b>	(en fil for Stavanger, 100m)
<b>-trondheim</b>	(en fil for Trondheim, 100m)
<b>-g2001</b>	(statistikk på ruter fra 1.1.2001 - ikke klar pt)
<b>-sasprogram</b>	(SAS-programmer for å lage rutestatistikk)
<b>-1_lage_ru1000.sas</b>	
<b>-1_lage_ru100.sas</b>	
<b>-2_agg_på_landet</b>	
<b>-2_agg_på_trondheim</b>	
<b>-2_agg_på_bergen</b>	
<b>-2_agg_på_stavanger</b>	
<b>-2_agg_på_oslo</b>	
<b>-2_agg_på_fylke</b>	
<b>-macro_var.sas</b>	
<b>-kart</b>	(rutenett som også legges ut på internett)
<b>-shape</b>	
<b>-landet</b>	(1000m)
<b>-oslo</b>	(100m)
<b>-bergen</b>	(100m)
<b>-stavanger</b>	(100m)
<b>-trondheim</b>	(100m)
<b>-SOSI</b>	
<b>-landet</b>	(1000m)
<b>-oslo</b>	(100m)
<b>-bergen</b>	(100m)
<b>-stavanger</b>	(100m)
<b>-trondheim</b>	(100m)

## Vedlegg B: Programmer for tilrettelegging av statistikken

```

/*****
Prosjekt .....: Rutenett
Skrevet av .....: Mona Takle og Gjermund Nygårdseter
Dato .....: 05.03.2002
Programmets funksjon ..: Lager rute-ID på individfilen; 1000 x 1000 m
Programmet heter .....: x:\320\1_regional\ruter\sasprogram\1_lage_rul000m.sas
Fil inn .....: $BESTAT/folkem/wk14/bosatte/gÅRET.sas7bdat
Fil ut .....: $BESTAT/folkem/wk14/bosatte/gÅRET.sas7bdat (oppdatert)
*****/
LIBNAME folk_bos '$BESTAT/folkem/wk14/bosatt';

DATA ruter;
SET folk_bos.g2000m01d01;
  IF oest_koo NE ' ' THEN DO;
    oest1=((INT(oest_koo/1000))*1000);
    nord1=((INT(nord_koo/1000))*1000);
    LENGTH oest $ 9;
    LENGTH nord $ 9;
    oest=oest1;
    nord=nord1;
    oest2=RIGHT(oest);
    nord2=LEFT(nord);
    rul000m=COMPRESS(oest2||nord2);
  END;

RUN;

DATA folk_bos.g2000m01d01;
SET ruter (DROP = oest nord oest1 nord1 oest2 nord2);
RUN;

/*****
Prosjekt .....: Rutenett
Skrevet av .....: Gjermund Nygårdseter
Dato .....: 12.09.2001
Programmets funksjon ..: Henter ut Landstall for rutenett.
.....: A. Trekker ut Landsstall fra situasjonsfil.
.....: B. Teller opp personer i ruter på aldersgrupper
Programmet heter .....:
x:\320\1_regional\ruter\sasprogram\2_agg_på_utvalgt_fylke_1000_meter.sas
Fil inn .....: $BESTAT/folkem/wk14/bosatte/gÅRET.sas7bdat
Fil ut .....: EXPORT: filen legges på
x:\320\1_region\ruter\gÅret\fNN\NN.txt
*****/
/*****
                                LES NØYE !!!!
NB NB NB      PROGRAMMET KJØRES MED SUBMIT !!!!! NB NB NB

1. VALG: Her velger du hvilket fylke du ønsker å kjøre for....
EKS: 04 = HEDMARK

2. FIL:   Her velger du hvilket situasjonsfil du ønsker å bruke.
EKS: g2000m01d01 = situasjonsfil pr. 01.01.2000
*****/

/* 1.*/ %let FYLKE = 04;
/* 2.*/ %let FIL   = g2000m01d01;
Run;

%include 'x:\320\1_regional\ruter\sasprogram\macro_var.sas';
run;
```

```

Rsubmit;
libname ruter '$REGION/rute/wk24';
libname sitfil '$BESTAT/folkem/wk14/bosatt';
run;

/*****
A. Trekker ut fylke fra situasjonsfilen.
*****/

Data land_ut;
SET sitfil.&fil;
WHERE oest_koo NE ' ' ;
RUN;

PROC SORT data = land_ut;
BY rul000m;
RUN;

/*****
B. Teller oppp personer i ruter på aldersgrupper
*****/

DATA Fil_til_kunde(drop=kjonn alderu);
SET land_ut (KEEP= rul000m kjonn alderu);
BY rul000m;
totalt+1;
IF kjonn='1' THEN menn+1;

IF kjonn='2' THEN kvinner+1;

IF alderu < '006' THEN T0_5+1;
IF '005' < alderu < '016' THEN T6_15+1;
IF '015' < alderu < '020' THEN T16_19+1;
IF '019' < alderu < '040' THEN T20_39+1;
IF '039' < alderu < '067' THEN T40_66+1;
IF '066' < alderu THEN T67+1;

IF LAST.rul000m THEN DO;
OUTPUT;
T0_5=0;
T6_15=0;
T16_19=0;
T20_39=0;
T40_66=0;
T67=0;
totalt=0;
menn=0;
kvinner=0;
END;

RUN;

Endrsubmit;

PROC EXPORT DATA= WORKUNIX.FIL_TIL_KUNDE
OUTFILE= "X:\320\1_regional\ruter\&aar\&landet\&utfil2"
DBMS=TAB REPLACE;

RUN;

```



## Vedlegg C: Program for å lage rutenettet i ArcInfo

```
/* *****
/* rutenett.AML
/*
/* MVD - 05.01.98
/*
/* AML som danner rutenett for en kommune.
/* Den tar utgangspunkt i et annet covers boundary-fil - .BND.
/*
/* Argumenter : inncover, utcover og cellestørrelse i meter*meter
/*
/* Forutsetninger: inncover må ligge i workspace pga at bnd-fila skal brukes i TABLES
/*
/* *****

&ARGS inncover utcover bredde hoyde

&IF [NULL %inncover%] &THEN
  &RETURN USAGE: &RUN rutenett <inncover> <utcover> <celle-bredde i m > <celle-hoyde i m>
&IF [LENGTH %inncover%] > 8 &THEN
  &RETURN inncover kan ikke ha mer enn 8 posisjoner i navnet

&IF [NULL %utcover%] &THEN
  &RETURN USAGE: &RUN rutenett <inncover> <utcover> <bredde> <hoyde>
&IF [LENGTH %utcover%] > 8 &THEN
  &RETURN utcover kan ikke ha mer enn 8 posisjoner i navnet

&IF [NULL %bredde%] &THEN
  &RETURN USAGE: &RUN rutenett <inncover> <utcover> <bredde> <hoyde>

&IF [NULL %hoyde%] &THEN
  &RETURN USAGE: &RUN rutenett <inncover> <utcover> <bredde> <hoyde>

&IF NOT [EXISTS %inncover% -cover] &THEN
  &RETURN inncoveret eksisterer ikke!

&IF [EXISTS %utcover% -cover] &THEN
  &DO
    &SETVAR overg := [RESPONSE 'Utcoveret finnes fra for ! Vil du skrive over ? (ja/nei)' nei]
    &IF %overg% <> ja &THEN
      &RETURN Velg nytt navn paa utcoveret.
    &ELSE
      kill %utcover%
  &END

/* Leser inn ekstrem-verdiene til coveret inn i variable
TABLES
SEL %inncover%.bnd
&SETVAR x_min := [SHOW RECORD 1 ITEM xmin]
&SETVAR y_min := [SHOW RECORD 1 ITEM ymin]
&SETVAR x_max := [SHOW RECORD 1 ITEM xmax]
&SETVAR y_max := [SHOW RECORD 1 ITEM ymax]
Q

/* For at origo for rutenettet alltid skal kunne passe inn i et UTM-kilometer-nett gjøres
/* følgende :
&SV x_uten_dec = [TRUNCATE %x_min%] /* fjerner decimaler for å få heltall
&SV y_uten_dec = [TRUNCATE %y_min%]

&SV lengd_xmin = [LENGTH %x_uten_dec%]
&SV lengd_ymin = [LENGTH %y_uten_dec%]

&SV start = %lengd_xmin% - 3 /* gjør om til ant. hele kilometer
```

```

&SV origo_x = [SUBSTR %x_min% 1 %start%]
&SV origo_x = %origo_x% * 1000      /* gjør om til meter

&SV start = %lengd_ymin% - 3      /* gjør om til ant. hele kilometer

&SV origo_y = [SUBSTR %y_min% 1 %start%]
&SV origo_y = %origo_y% * 1000      /* gjør om til meter

/* Generering av rutenett:
&SV y_koor = %origo_y% + 500      /* legger 500 meter til y-aksen

GENERATE %outcover%
FISHNET
%origo_x%, %origo_y%
%origo_x%, %y_koor%
%bredde%, %hoyde%
0,0
%x_max%, %y_max%
Q

/* bygger polygontopologi:

BUILD %outcover%

&RETURN

```

## De sist utgitte publikasjonene i serien Notater

- 2001/73 L. Vågane: Omnibusundersøkelsen 2000. Dokumentasjonsrapport. 115s.
- 2001/74 A.S. Abrahamsen, G. Olsen: Bedrifts-  
populasjonen 1998. Registrering av nye og  
opphørte bedrifter. 58s.
- 2001/75 A. Wethal: Omlegging av produksjons-  
rutinene for statistikk over veitrafikk-  
ulykker. Dokumentasjon av prosessen og  
nytt produksjonsopplegg. 30s.
- 2001/76 B. Lie og G. Daugstad : Sammenlignende  
studie av norsk og finsk kultursektor. 175s.
- 2001/77 G. Haakonsen: Beregninger av utslipp til  
luft av klimagasser. En gjennomgang av  
arbeidsprosess og dokumentasjon. 39s.
- 2002/1 P. Scøning: Statistikk for 16 tettsteder og  
deres sentrumsarealer. Et innspill til  
programmet for utvikling av miljøvennlige  
og attraktive tettsteder i distriktene. 58s.
- 2002/2 V.V. Holst Bloch:  
Arealbruksklassifisering av bebygde  
arealer. Revidert rutine for tilordning av  
arealbruksklasse til bygning. 58s.
- 2002/3 V.V. Holst Bloch: Metode og  
datagrunnlag for produksjon av  
arealstatistikk for tett-stednære områder.  
Teknisk dokumenta-sjon. 29s.
- 2002/4 G. Dahl og J Lajord: FD - Trygd:  
Konsistenskontroller. 21s.
- 2002/5 G. Dahl og C. Nordseth: FD - Trygd:  
Dokumentasjonsrapport. Inntekt og  
formue, 1992-1999. 41s.
- 2002/6 G. Dahl og C. Nordseth: FD - Trygd:  
Dokumentasjonsrapport. Sysselsetting.  
1992-1998. 107s.
- 2002/7 S. Lien og C. Nordseth: FD - Trygd:  
Dokumentasjonsrapport. Pensjonsgivende  
inntekt, 1992-1999. Omsorgspoeng,  
1992-1997. 24s.
- 2002/8 V. Lund: Kostnadsindekser for  
lastebiltransport. Definisjoner og  
beregningsmetode. Vekter og  
representantvarer 2001. 47s.
- 2002/9 T. M. Normann: Rekruttering til  
erfaringskonferanse og undersøkelse om  
røykevaner blant kvinner i alderen 25-45  
år. Dokumentasjonsrapport. 16s.
- 2002/10 J. Holmøy: GERIX 1995-1999. Doku-  
mentasjon, system, data, program. 47s.
- 2002/11 T.M. Normann: Underøking om det lokale  
sjølvstyret. Dokumentasjonsrapport. 81S.
- 2002/12 L.S. Stambøl: Regionale framskrivinger av  
sysselsetting og bruttoprodukt ved hjelp av  
SSBs modellsystem REGARD. Regionale  
framskrivinger basert på nasjonale anslag  
med modellene MODAG (1997-2005) og  
MSG (1995-2020). 35s.
- 2002/13 H. Madsen og L S. Stambøl: Kontra-  
faktiske regionale beregninger ved hjelp  
av SSBs modellsystem REGARD.  
Regionale beregninger basert på historiske  
tall på nasjonalt nivå kjørt bakover i tid på  
grunnlag av modellens basisår (her 1995).  
55s.
- 2002/14 V. Hansen og H. Madsen: Månedlig og  
kvartalsvis elektrisitetsstatistikk. Doku-  
mentasjon av produksjonsrutiner og  
systembeskrivelse. 41s.
- 2002/15 A. Rolland: Søkelys på det gode liv. 37s.
- 2002/17 D.Rønningen og D. Fredriksen: Beskat-  
ningen av pensjonister. 41s.
- 2002/18 D. Rønningen: Overganger fra arbeids-  
markedet til trygd. En litteraturoversikt.  
34s.
- 2002/19 F. Gundersen og L. Solheim:  
Regionalisering av FoU-statistikken. 43.
- 2002/20 L. Vågane: Omnibusundersøkelsen  
november/desember 2001.  
Dokumentasjonsrapport. 56s.
- 2002/21 G. Claus, O. Haugen P. M. Holt  
E. Knutsen: Regnskapsstatistikk.  
Næringsoppgaver for ikke-finansielle  
aksjeselskaper, 1999. Dokumentasjon.  
34s.