

Aslaug Hurlen Foss

Notater

Kvaliteten i arbeidsmarkedsdelen i Folke- og boligtellingsen 2001

Innhold

1. Innledning	3
1.1. Formålet	3
1.2. Oppsummering av resultatene	3
1.3. Innholdet	4
2. Beskrivelse av datasettene og koplingen mellom de	5
2.1. Populasjonen i arbeidsmarkedsstatistikk	5
2.2. Arbeidskraftundersøkelsen, AKU	5
2.3. Registerbasert sysselsetningsstatistikk. Arbeidsmarkedsregisteret	5
2.3. Koplingen mellom arbeidsmarkedsregisteret og AKU	6
3. utfordringer med å sammenligne undersøkelser	7
3.1. Definisjoner, presiseringer og kilder for variablene	7
3.2. Feilkilder i arbeidsmarkedsregisteret	10
3.3. Feilkilder i AKU	11
3.4. Forskjellige tidspunkter for undersøkelsen	13
3.5. Analyse av mulige feilkilder i AKU og arbeidsmarkedsregisteret	13
3.6. Konsekvenser av bruk av indirekte intervju i AKU	16
4. Målefeil og utvalgsfeil	19
5. Analyse av kvaliteten	22
5.1. Sysselsatt	22
5.1.1. Kvaliteten av kildedataene	23
5.1.2. Andelen sysselsatte etter alder	24
5.1.3. Andelen sysselsatt etter fylke	25
5.1.4. Andelen sysselsatt etter utdanning	27
5.2. Yrkesstatus	29
5.3. Aktivitetsstatus	30
5.4. Sysselsatte etter næring	32
5.5. Avtalt/vanlig (gjennomsnittlig) arbeidstid per uke	34
5.6. Sektor	36
5.7. Antall sysselsatte i bedriften	37
Referanser	40
De sist utgitte publikasjonene i serien Notater	41

1. Innledning

1.1. Formålet

Dette notatet beskriver kvaliteten i registerbasert sysselsetningsstatistikk (heretter kalt arbeidsmarkedsregister) som er publisert som en del av Folke- og bolig tellingen i 2001. Dette blir gjort ved å sammenligne opplysninger fra arbeidsmarkedsregisteret med arbeidskraftundersøkelsen, heretter kalt AKU. Det vil bli gjort en sammenligning av disse to undersøkelsene både på makro og mikro nivå. Koplingen mellom disse to undersøkelsene blir gjort ved å kople sammen filene på personnivå, det vil si ved hjelp av fødselsnummer. Sysselsetningsstatus er direkte knyttet til personen, mens for de andre variabler næring, sektor, arbeidstid, yrkesstatus, aktivitetsstatus og antall sysselsatte i bedriften er knyttet til arbeidsforhold og ikke person. I arbeidsmarkedsregisteret og AKU kan det være valgt forskjellig hovedarbeidsforhold, som kan føre til forskjeller. Det er også andre store utfordringer ved en slik sammenligning. For det første er undersøkelsene fra forskjellig tidspunkt. Opplysninger om sysselsatte i arbeidsmarkedsregisteret er fra tellingsuka, 29. oktober - 4. november 2001, mens AKU gir tall for alle ukene i 4. kvartal 2001, det vil si uke 40 til 52. Selv om det kan være en del uker mellom tidspunktene er antakelig endringen i sysselsetningstallene små, slik at det har liten betydning. En annen utfordring er forskjellige definisjoner og produksjonsprosess av variabler. Definisjonene som er brukt i arbeidsmarkedsregisteret er i hovedsak lik AKU. Produksjonsmetoden for variablene er derimot forskjellig og dette vil være en kilde til å forstå forskjellene som oppstår i analysen. En annen utfordringen er at det forekommer feil i både arbeidsmarkedsregisteret og i AKU, men de er forskjellige. I AKU vil kvaliteten hovedsakelig avhenge av utvalgsstørrelse, frafall, hvor nøyaktig oppgavegiveren svarer på spørsmål og hvordan svarene blir kodet. Mens kvaliteten i arbeidsmarkedsregisteret vil hovedsakelig avhenge av kvaliteten i kilderegistrene og de metodiske valgene som er tatt i produksjonsprosessen.

1.2. Oppsummering av resultatene

Kvaliteten i arbeidsmarkedsregisteret blir analysert ved å sammenligne data fra registeret med AKU. Dette gjøres både for total tall og på individnivå. I de fleste tilfeller vil AKU bli antatt å ha den sanne verdien, selv om vi vet at det fins feil i AKU. Den feilen vi måler i registerstatistikken er slik sett iforhold til hva en totaltelling med AKU ville ha gitt. Arbeidsmarkedsregisteret er bygd opp gjennom koplingen mellom mange forskjellige datakilder. De viktigste kildene er arbeidstakerregisteret, lønns- og trekkoppgaveregisteret, selvangivelsesregisteret og SOFA-søkerregisteret. Den beste og største datakilden er koplingen mellom arbeidstakerregisteret og lønns- og trekkoppgaveregisteret.

- Kvaliteten på andelen sysselsatte er forholdsvis god ved sammenligning med AKU. Dette gjelder både når tallene er sammenlignet på makro og mikro nivå. I arbeidsmarkedsregisteret blir det fylt opp med sysselsatte til nivået blir likt med AKU.
- Andelen sysselsatte i følgende grupper har god kvalitet:
 - kjønn
 - fylke
 - utdanning
 - næring (inndeling i 7 hovedgrupper)
- For andelen sysselsatte for de i alderen 25-66 år er det en god kvalitet, men for de mellom 20-24 år og 67-74 år er det relativt store systematiske skjevheter. Grunnen til dette er at det er vanskelig å klassifisere personer med deltidsarbeid eller personer med sporadiske arbeidsforhold riktig i arbeidsmarkedsregisteret.

- For andelen sysselsatte etter avtalt/vanlig arbeidstid er det et forholdsvis godt samsvar på makro nivå, men på mikro nivå er det bare et godt samsvar for de som arbeidet mer enn 30 timer.
- Andelen arbeidsledige, under utdanning, pensjonister og andre ikke sysselsatte er det er klart svakere samsvar for mellom arbeidsmarkedsregisteret og AKU.
- Ifølge arbeidsmarkedsregisteret er det færre ansatte i statlig og kommunal/fylkeskommunal sektor enn ifølge AKU. Grunnen til denne forskjellen kan være valg av hovedarbeid eller feil koding av sektor, spesielt for offentlig foretak.
- Samsvaret mellom AKU og registeret for antall sysselsatte i bedriften er svært dårlig. Dette skyldes antakelig at det er et vanskelig spørsmål å svare på, for de som er med i AKU.

1.3. Innholdet

I kapitel 2 er datasettene som er benyttet i analysen beskrevet. I tillegg er det beskrevet hvordan datasettene er koplet sammen. En utfordring er forskjellen i definisjoner eller konsepter for variabler som er benyttet i disse to undersøkelsene. Dette er nærmere behandlet i kapitel 3. I tillegg er feilkildene i AKU og arbeidsmarkedsregisteret beskrevet og analyser i dette kapitlet. I kapitel 4 er de statistiske målene for å estimering av kvaliteten beskrevet. Til slutt i kapitel 5 er kvaliteten til arbeidsmarkedsopplysningene i Folke- og bolig tellingen analysert.

2. Beskrivelse av datasettene og koplingen mellom de

2.1. Populasjonen i arbeidsmarkedsstatistikk

I arbeidsmarkedsstatistikk er populasjonen alle som er bosatte i Norge i alderen 16 til 74 år. Det vil si alle personer som er i en alder der det er normalt at man kan arbeide. Det er folkeregisteret som holder rede for hvem som er bosatt i Norge. SSB har en kopi av folkeregisteret som er kaldt Bebas, denne basen blir fortløpende oppdatert med endringer i populasjonen. I tillegg er denne basen utvidet med blant annet informasjon om hvem som er innvandrere.

2.2. Arbeidskraftsundersøkelsen, AKU

AKU er en utvalgsundersøkelse. I AKU blir det trukket et tilfeldig utvalg av familieenheter fra Bebas. Familietypene er bare de offisielle familietypene fra DSF: Enslig, gift, gift med barn, Far/mor med barn og andre. Utvalget består av til sammen 24 000 personer (per kvartal) i aldersgruppa 16-74 år. Hvert familiemedlem i denne alderen intervjues om sin tilknytning til arbeidsmarkedet i en nærmere spesifisert referanseuke. Intervjuobjektene deltar i 8 påfølgende kvartaler.

Intervjuet blir gjort hovedsakelig per telefon eller eventuelt ved besøk hvis ikke det blir oppnådd telefonkontakt. En del opplysninger gitt i tidligere intervju brukes ved etterfølgende intervju. For koding av næring hentes det inn informasjon fra Arbeidstakerregisteret og Det sentrale Bedrifts- og Foretaksregisteret (på individnivå). For øvrig innhentes demografiske data fra personregisteret Bebas, og utdanningsdata fra SSBs register over befolkningens høyeste utdanning.

Oppgavegiver er i utgangspunktet samme som observasjonsenheten. I noen tilfeller blir informasjonen samlet inn gjennom indirekte intervju, dvs. ved svar fra nære familiemedlemmer. I 4. kvartal 2001 gjaldt dette 14 prosent.

De absolutte tallene fra AKU presenteres i form av estimerte totaltall for hele befolkningen i alderen 16-74 år. Oppblåsningsfaktoren varierer, men har et gjennomsnitt på om lag 140. Faktorene beregnes for 109 strata basert på demografiske data og registeropplysninger om tilknytningen til arbeidsmarkedet.

Svarprosenten er om lag 90 prosent. Det justeres for totalfrafallet ved estimering. Partielle frafall imputeres for enkelte sentrale variabler etter et sett med regler (maskinell imputering). For en nærmere dokumentasjon, se Bø og Håland (2002).

2.3. Registerbasert sysselsetningsstatistikk. Arbeidsmarkedsregisteret

Data for arbeidsmarkedet i tellingen er basert på flere ulike registre. Utgangspunktet er alle personer som er bosatt i Norge per 3. november 2001 ifølge folkeregisteret. Det er bare tatt med de personene som er i alderen 16-74 år ved utgangen av året slik som i AKU. Populasjonen er i hovedsakelig lik slik den er i AKU bortsett fra at de har trukket utvalget fra populasjonen på et tidligere tidspunkt. De viktigste er registrene for opplysninger om arbeidstilknytting er arbeidstakerregisteret, lønns- og trekkoppgaveregisteret (LTO), selvangivelsesregisteret, SOFA-søkerregisteret og Enhetsregisteret/Bedrifts- og foretaksregisteret (BoF). Arbeidstakerregisteret er hovedkilden til data om lønnstakere, men LTO utgjør et viktig supplement. Begge registre har arbeidsforhold (jobber) som enhet. I arbeidstakerregister er det mulig å være registrert med flere jobber i samme foretak, mens dette i LTO kommer bare som lønn fra en jobb i et foretak. Selvangivelsesregisteret er hovedkilden til

opplysninger om selvstendig næringsdrivende. Enhetsregisteret og BoF gir opplysninger om bedriftene (arbeidsstedene), mens SOFA-søkerregisteret gir data om arbeidsledige og personer på arbeidsmarkedstiltak. I tillegg nyttes supplerende data fra en rekke andre kilder: Registre over vernepliktige og personer i sivilteneste, registre over ansatte i stat og kommune, lønnsstatistikk for ansatte i privat sektor, sykefraværsregisteret mv.

For personer som er bosatt i Norge, men som jobber i et annet land vil arbeidsmarkedsregisteret ikke ha noen opplysninger om arbeidsforholdet, siden det ikke finnes noe informasjon om dette i de norske kilderegistrene. Dette vil være en forskjell til AKU hvor det ikke er noen begrensning på hvor personen jobber. For kommuner med høy andel av personer som jobber i et annet land vil derfor andelen sysselsatte være for lav.

Opplysninger om arbeidsmarked i tellingen er altså basert på en rekke ulike kilder, og det er bygget opp et system i SSB for en samlet utnyttning av disse. Systemene omfatter moduler for konsistensbehandling mellom ulike datakilder, valg av viktigste arbeidsforhold og klassifisering som sysselsatt. Dokumentasjon av det nåværende systemet er under utarbeidelse. For dokumentasjon av tidligere klassifiseringsrutiner, se Bråten og Fosen 1998. For valg av klassifikasjonsrutiner se, Fosen og Solheim 2000.

2.3. Koplingen mellom arbeidsmarkedsregisteret og AKU

Koplingen mellom AKU og arbeidsmarkedsregisteret ble gjort ved hjelp av fødselsnummer. Dette vil si at opplysninger fra registeret ble koplet på AKU-filen på personnivå. For sysselsetningsstatus blir koplingen på person det mest korrekte, men for variabler som gjelder arbeidsforhold vil ikke alltid denne koplingen være korrekt. Det kan være valgt forskjellige arbeidsforhold som det viktigste i arbeidsmarkedsregisteret og AKU. Det vil altså bli en analyse av forskjellen av det viktigste arbeidsforholdet i arbeidsmarkedsregisteret og AKU og ikke av selve arbeidsforholdet. I analysen vil vektene i AKU bli benyttet i beregningen av målefeil. Analysen vil vise forskjellen mellom om vi brukte registervariabler og AKU-variabler på AKU-utvalget. På grunn av skjevheter i utvalget kan estimatene fra AKU-utvalget med registervariabler avvike noe fra resultatene fra arbeidsmarkedsregisteret.

3. utfordringer med å sammenligne undersøkelser

Det er mange utfordringer med å sammenligne undersøkelser. For det første forekommer det feil i begge kilder som kan påvirke resultatene. For det andre kan det være forskjellige definisjoner og konsepter som er benyttet i to forskjellige undersøkelser.

3.1. Definisjoner, presiseringer og kilder for variablene

Forskjellige i definisjoner eller hvordan variablene er laget vil være en årsak til målefeilen. I AKU og arbeidsmarkedsregisteret er det forsøkt å bruke mest mulig like definisjoner, men det kan være store forskjeller på hvordan de er laget. I tabell 3.1.1 er de viktigste variablene med definisjoner listet opp.

Tabell 3.1.1. Definisjoner, presiseringer og kilder for variabler i AKU og arbeidsmarkedsregisteret

Variabel		
Sysselsatte	Definisjon	Sysselsatte er definert som personer som utførte inntektsgivende arbeid av minst én times varighet i referanseuka, samt personer som har et slikt arbeid, men som var midlertidig fraværende pga. sykdom, ferie, lønnet permisjon e.l. Personer som er inne til førstegangs militær- eller sivil tjeneste regnes som sysselsatte. Personer på sysselsettingstiltak med lønn fra arbeidsgiver klassifiseres også som sysselsatte.
	AKU	I AKU blir definisjonen fulgt og sysselsatte blir bestemt ut fra spørsmål om arbeidsforhold og antall inntektsgivende timer.
	Register	I tellingen og annen registerbasert statistikk er målet ved hjelp tilgjengelige data å lage en klassifisering av sysselsatte som ligger så nær opp til definisjon som mulig. Det viktigste registeret for å identifisere sysselsatte er arbeidstakerregisteret som har dateringer som viser hvilke arbeidsforhold som var aktive i referanseuka (tellingsuka). Kriteriene for å bli registrert i arbeidstakerregisteret er: 'Arbeidsforhold med minst fire timers gjennomsnittlig arbeidstid per uke. Det er videre et kriterium at arbeidsforholdet skal vare i mer enn seks dager'. De andre hoveddatakildene, lønns- og trekkoppgaverregisteret og selvangivelsesregisteret har ikke en slik datering. Hvilke personer som skal klassifiseres som sysselsatte i referanseuka er da beregnet ut fra opplysninger om lønn og næringsinntekt, samt supplerende informasjon fra andre registre og statistikkssystemer. Metoden gir et tall på sysselsatte i alt som stemmer overens med det tilsvarende tall fra AKU. For sysselsatte med flere arbeidsforhold i referanseuka, fastsettes ett som det viktigste.
	Forskjeller	Det er ingen definisjonsforskjeller, men bare i hvordan variablene er laget. Hovedforskjellen mellom AKU og arbeidsmarkedsregisteret er at i AKU er det personen selv som har rapporter om arbeidsforholdet, mens i registeret er det rapportert inn fra bedriften eller imputert.

Variabel		
Yrkesstatus	Definisjon	Denne variabelen skiller mellom lønnstakere og selvstendige og referer seg til det viktigste arbeidsforholdet. <i>Lønnstakere</i> , eller ansatte, omfatter også medeiere i aksjeselskaper og andre selskaper med begrenset ansvar. <i>Selvstendige</i> er personer som driver næringsvirksomhet for egen regning (alene eller sammen med andre). Det finnes også en tredje gruppe, <i>familiearbeidere</i> (familiemedlemmer uten fast lønn i familiebedrift).
	AKU	Definisjonen blir fulgt og opplysningene kommer fram ved spørsmål i intervjuet.
	Register	Definisjonen blir fulgt, utenom at det ikke blir gitt tall for familiearbeidere, da mange av disse ikke er mulig å identifisere i registrene. Antall sysselsatte ifølge registrene og da fordelt etter lønnstakere og selvstendige, tilsvarer antallet gitt i AKU. Klassifiseringen av selvstendige næringsdrivende er basert på informasjon om næringsinntekt fra Selvangivelsesregisteret fra året før, samt opplysninger om andre aktive lønnstakerforhold på referansetidspunktet.
	Forskjeller	I AKU er selvstendige tatt direkte ut fra spørsmålene, mens i arbeidsmarkedsregisteret er det beregnet ut fra lønnstakerforhold og næringsinntekt. Det blir ikke gitt tall for familiearbeidere i arbeidsmarkedsregisteret, men i AKU.
Næring	Definisjon	Næring beskriver typen av produksjon eller aktiviteter på arbeidsstedet (bedriften). Næring blir kodet etter Standard for næringsgruppering fra 1994 (NACE), se NOS C182.
	AKU	Definisjonen blir fulgt i AKU. Koding av næring blir gjort ut fra spørsmål i intervjuet, og i tillegg blir Bedrifts- og Foretaksregisteret (BoF) brukt som en støtteinformasjon. I 4. kvartal 2001 var det blant de sysselsatte 50 prosent som hadde samme bedrift som i BoF, 9 prosent samme for foretak, 17 prosent ikke samsvar mellom AKU og BoF og 23 prosent hadde ingen informasjon i BoF.
	Register	Definisjonen blir også fulgt i registeret. Næring er innhentet fra bedrifts- og foretaksregisteret, og gjelder bedriften hvor personen arbeider. For selvstendig næringsdrivende som ikke kan knyttes til verken bedrift eller foretak, ligger informasjon om utdanning og familiemedlemmers selvstendige virksomhet til grunn.
	Forskjeller	Både AKU og arbeidsmarkedsregisteret bruker BoF som kilde, men i AKU går informasjon fra intervju forran hvis det er forskjeller. For de som ikke har opplysninger om næring, vil næring bli kodet ut fra spørsmål i AKU eller andre registeropplysninger i arbeidsmarkedsregisteret. For de med flere jobber kan det bli forskjell ut fra hva som er kodet som hovedjobb.

Variabel		
Sektor	Definisjon	Sektor tar utgangspunkt i <i>Institusjonell sektorgruppering</i> som er en statistisk standard basert på anbefalinger for nasjonalregnskapsoppstilling gitt av FN (SNA 93), se Nasjonalregnskapsstatistikk 1992-1999 (NOS C613). Det er benyttet en hovedinndelingen i fire grupper: <i>Statlig forvaltning, fylkeskommunal forvaltning, kommunal forvaltning</i> og <i>privat sektor og offentlige foretak</i>
	AKU	Informasjon om sektor er i AKU hentet inn fra intervju. Informasjonen blir imidlertid kontrollert mot bedriften for å oppdage åpenbare feil.
	Register	Sektor blir laget ut fra hvilket foretak eller næring hver person er registrert på. Privat sektor og offentlig foretak omfatter foretakssektoren, husholdningssektoren og ideelle organisasjoner samt sysselsatte med uoppgitt sektor. For foretak som skifter kategori kan det ta tid før dette blir registrert.
	Forskjeller	Det er ingen forskjell mellom definisjonene i AKU og register. Det kan likevel bli forskjeller fordi opplysningene er framkommet på forskjellig måte. For offentlige foretak som blir fristilt kan det være en gråsoner for hvilke foretak som offentlig og hvilke som er privat. Dette problemet gjelder både AKU og arbeidsmarkedsregisteret.
Avtalt/vanlig (gjennomsnittlig) arbeidstid per uke	Definisjon	Avtalt arbeidstid er det antall arbeidstimer pr. uke som den ansatte ifølge arbeidskontrakten skal være på arbeid. Eventuelt fravær fra arbeidet pga. sykdom, ferie e.l. skal ikke trekkes fra i avtalt arbeidstid, og overtid skal ikke medregnes.
	AKU	For ansatte med avtale om arbeidstid hvor det avtalte timetallet varierer fra uke til uke, f.eks. som en følge av skift- eller turnusordning, oppgis både den arbeidstida som gjaldt for selve undersøkelsesuka og gjennomsnittet av den avtalte arbeidstida pr. uke (i tabeller over avtalt arbeidstid i denne publikasjonen er det gjennomsnittstallene som brukes). For ansatte som ikke har noen arbeidstidsavtale, samt for selvstendige og familiearbeidere, får vi oppgitt deres vanlige (gjennomsnittlige) arbeidstid pr. uke (normalt for de siste 4 ukene).
	Register	For lønnstakere er dette <i>avtalt</i> arbeidstid i uka i det viktigste arbeidsforholdet, dvs. det antall arbeidstimer per uke som den ansatte ifølge arbeidskontrakten skal være på arbeid. For selvstendige er det beregnet <i>vanlig</i> arbeidstid per uke ut fra tilgjengelige registeropplysninger. For 90 prosent av lønnstakerene innhentes informasjonen fra arbeidstakerregisteret. For personer klassifisert som lønnstakere på grunnlag av informasjon fra lønns- og trekkoppgaverregisteret, beregnes <i>arbeidstid</i> ut fra varigheten av lønnsforholdet og størrelsen på lønn. <i>Arbeidstid</i> for selvstendige vil være gjennomsnittlig arbeidstid per uke, beregnet på grunnlag av næringsinntekt og informasjon fra Arbeidskraftundersøkelsen.
	Forskjeller	I arbeidsmarkedsregisteret blir det for de som ikke er registrert i arbeidstakerregisteret imputert hvor mange timer som er jobbet, mens det i AKU er kommet ut fra et spørsmål til hver enkelt person.

Variabel		
Antall sysselsatte i bedriften	Definisjon	Ingen spesiell definisjon. Kriteriene for klassifisering av sysselsatt må være fulgt.
	AKU	I AKU er informasjonen hentet fra spørsmål til intervjuobjekter og svarene er oppgitt i intervaller
	Register	Opplysningene om antall sysselsatte i bedriften er laget i hovedsak ut fra optelling av hvor mange personer som er registrert som sysselsatt i en bedrift ifølge AA-registeret. Det er i tillegg lagt til de som er sysselsatt ifølge LTO og har et koplet bedriftsforhold. Bedriftens størrelse er målt i antall sysselsatte ved bedriften i tellingsuka. Alle sysselsatte er regnet med, også de som har sitt viktigste arbeidsforhold i en annen bedrift.
	Forskjeller	Forskjellen er hovedsakelig hvordan dataene er samlet inn og at AKU bare har intervaller.
Utdanning	Definisjon	Befolkningens høyeste utdanning. Utdanningene er gruppert etter Norsk standard for utdanningsgruppering (NUS) som første gang ble opprettet av SSB i 1970 og ble sist oppdatert i 2000 (NUS2000).
	AKU	Ved første intervju blir opplysninger om IO's høyeste utdanning hentet fra register. Opplysningene blir kontrollert med et spørsmål i tilfelle IO har tatt mer utdanning. Hvis utdanningen er uoppgitt i registeret blir IO spurt om utdanningen og den blir deretter kodet etter den reviderte standarden fra 2000. Ved senere intervjuer blir IO spurt om han/hun har tatt mer utdanning.
	Register	Opplysningene i tellingen blir hentet fra registeret BHU (befolkningens høyeste utdanning) per 1. oktober 2001.
	Forskjeller	I AKU blir opplysningene fortløpende oppdatert med opplysninger fra IO. I tillegg er det færre med uoppgitt utdanning i AKU, siden disse får spørsmål om sin utdanning som blir kodet.

3.2. Feilkilder i arbeidsmarkedsregisteret

Forsinkelse av meldinger

Oppdateringen av arbeidstakerregisteret, som er hovedkilden i den registerbaserte sysselsetningsstatistikken, vil alltid ligge litt på etterskudd i forhold til virkeligheten. Det vil ta litt tid fra et nytt arbeidsforhold er inngått før det er registrert. Før statistikken blir tatt ut blir det derfor ventet 4 måneder på etterslep av meldinger.

Mangelfull datering

For personer som er definert som sysselsatte og lønnstakere kun på grunnlag av opplysninger fra lønns- og trekkoppgaveregisteret (omkring 10 prosent av lønnstakerne), er arbeidsforholdet ikke datert. For rundt halvparten av denne gruppen innhentes informasjon fra andre administrative kilder, som er med på å tidfeste arbeidsforholdet. For de resterende legges informasjon om lønn til grunn for om en person anses som sysselsatt. Det er dermed knyttet en viss usikkerhet om arbeidsforholdet faktisk var aktivt i referanseuken.

Lang produksjonstid

Selvstendig næringsdrivende identifiseres ved hjelp av informasjon fra Selvangivelsesregisteret. Grunnet lang produksjonstid ligger opplysninger om næringsvirksomhet fra året før til grunn for

utarbeidelse av statistikken. Som følge av denne tidsforskyvningen kan personer dermed feilaktig bli klassifisert som sysselsatte, hvis de avsluttet virksomheten det foregående året. Det motsatte problemet gjelder for de som starter opp som selvstendig næringsdrivende, siden de ikke vil bli registrert før det er gått et år. Det vil altså bli en del feil i hvilke personer som er selvstendige, mens nivået for andelen selvstendige kan være riktig.

Identifisering av bedriften

For personer som er definert som sysselsatte og lønnstakere kun på grunnlag av opplysninger fra lønns- og trekkoppgaveregisteret er arbeidsforholdet knyttet til et foretak. Her er det utarbeidet en rutine for på best mulig måte identifisere bedriften. I de tilfeller hvor personen er ansatt i et flerbedriftsforetak, kan det være usikkert om arbeidsforholdet faktisk blir knyttet til riktig bedrift, og dermed får korrekte opplysninger om næring og arbeidssted. For store foretak med mange bedrifter under seg, er fordelingen av ansatte i arbeidstakerregisteret til tider mangelfull. Det kan gi merkbare utslag på kommunenivå når slike feil oppstår og når de rettes.

3.3. Feilkilder i AKU

Utvalgsfeil

På grunn av det er bare en liten del av populasjonen som blir intervjuet vil det være en del usikkerhet i estimatene. Selv om utvalget er representativt i forhold til viktige variabler som alder og kjønn er det ikke sikkert at utvalget er representativt i forhold til det som skal undersøkes. Denne usikkerheten kaller vi utvalgsfeil.

Målefeil

De fleste opplysningene i AKU fremkommer ved intervju. Det vil i disse intervjudataene være en del målefeil, det vil si at personene ikke svarer korrekt på spørsmålene som blir stilt. Grunnen til dette kan være fordi de ikke er sikker på svaret, ikke har skjønt spørsmålet riktig eller ikke ønsker å svare riktig.

Indirekte intervju

Fjerde kvartal 2001 var det 14 prosent av intervjuene som var indirekte. Det vil si at en i familien svarte istedenfor selve intervjuobjektet. Dette kan ha ført til feilaktive opplysninger i AKU da det ikke er sikkert at alle arbeidsforhold var kjent for den som svarte. I tabell 3.3.1. ser vi at samsvaret i sysselsetning er større mellom AKU og register når det er gjort et direkte intervju i forhold til et indirekte intervju. En mer utfyllende analyse av dette problemet kommer i avsnitt 3.5 og 3.6.

Tabell 3.3.1. Samsvar i sysselsetning for direkte og indirekte intervju. AKU og register

	Direkte intervju	Indirekte intervju
Sysselsatt i både AKU og register	12 612	1 686
Sysselsatt i AKU i alt	13 361	1 821
Prosent samsvar	94,4	92,6

Frafall

I en utvalgsundersøkelse vil det alltid være noen som ikke svarer på undersøkelsen. I AKU utvalget 4. kvartal 2001 var det trukket ut 24 258 personer. På tellingstidspunktet var 254 av disse personene ikke lengre bosatte i Norge. I tillegg var det 2 788 personer som ikke svarte på undersøkelsen. Dette frafallet kan ha ført til at utvalget ikke blir representativt for hele populasjonen. Dette gjelder særlig hvis en gruppe blir underrepresentert i forhold til de andre. En del av problemene med frafall blir kompensert for ved hjelp av vekting. I tabell 3.3.2. blir noen registervariabler vist for AKU både med og uten vekter og for frafallet. Disse blir sammenlignet med tallene fra tellingen. Av tabellen ser vi at frafallet er størst for menn, personer under 40, de som bor alene, for de med lav utdanning og for innvandrere. Andelen sysselsatte ifølge arbeidsmarkedsregisteret er lavere i frafallet enn blant de som har svart i AKU. Det er 73 prosent sysselsatt i AKU mot 64 prosent i frafallet til AKU. Andelen

sysselsatte i privat sektor er høyere i frafallet enn blant de som har svart i AKU. For variabelen næring er det et forholdsvis stort frafall blant de i varehandel, hotell- og restaurantvirksomhet.

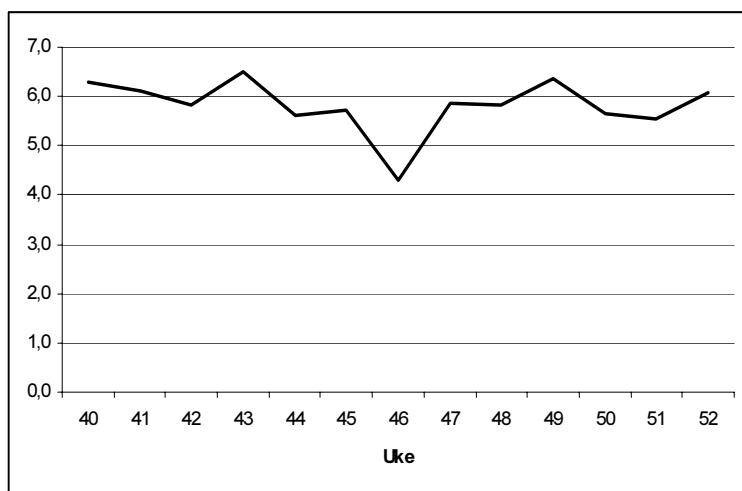
Tabell 3.3.2. Tellingen, AKU, AKU uten vektorer og frafallet. Registervariabler. Prosent

Register variabler	Telling (16-74 år)	AKU	AKU uten vektorer	AKU, fracfall	Frafall i prosent av utvalget
Andelen menn	50	50	50	55	13
Alder					
I alt	100	100	100	100	12
16-19 år	7	7	7	5	8
20-24 år	8	8	8	12	17
25-39 år	31	31	30	40	15
40-54 år	29	29	30	26	10
55-66 år	17	16	17	13	9
67-74 år	8	8	8	3	4
Utdanning					
I alt	100	100	100	100	12
Grunnskolenivå	17	17	17	18	12
Videregående skole nivå	57	56	57	55	11
Univeristet og høyskolenivå	23	25	24	20	10
Uoppgitt eller ingen fullført utdanning	3	2	2	7	36
Innvandrerkategori					
I alt	100	100	100	100	12
Uten innvandringsbakgrunn	89	90	91	80	10
Førstegenerasjonsinnvandrere	7	5	5	15	28
Annet	4	5	4	5	13
Husholdningsstørrelse					
I alt	100	100	100	100	12
1 person	18	16	15	35	23
2 personer	27	27	27	23	10
3 personer	20	21	21	16	9
4 personer	21	22	23	15	8
5 personer eller flere	14	14	15	10	8
Andelen sysselsatt	70,8	71,5	72,8	63,7	10
Andelen ikke sysselsatt	29,2	28,6	27,2	36,3	15
Andelen av sysselsatte i privatsektor	70,7	70,3	70,0	75,8	13
Sysselsatt etter næring					
I alt	100,0	100,0	100,0	100,0	12
01-05 Jordbruk, skogbruk og fiske	3,7	3,9	4,1	3,0	8
10-37 Utvinning av råolje og naturgass, industri og bergv.	15,1	14,9	15,5	14,4	10
40-45 Bygge- og anleggsvirksomhet, kraft- og vannf.	6,7	6,7	7,0	6,0	9
50-55 Varehandel, hotell- og restaurantvirksomhet	18,2	17,6	17,4	22,9	13
60-64 Transport og kommunikasjon	7,3	7,0	6,9	8,7	13
65-74 Finansiell og forretningsmessig tjenesteyting, eiend.	12,4	12,6	11,7	12,8	11
75-99 Offentlig administrasjon og forsvar, helse- og sosialt.	35,9	36,8	36,9	30,9	9
Uoppgitt	0,6	0,5	0,5	1,2	23

3.4. Forskjellige tidspunkter for undersøkelsen

Statistikken fra AKU gjelder for alle ukene i 4. kvartal 2001, mens statistikken fra arbeidsmarkedsregister bare gjelder for uke 44, det vil si fra 29. oktober til 4. november. I tabell 3.4.1. er det vist den prosentvise feilen: sysselsatt i AKU, men ikke i registeret av totalt antall sysselsatte i AKU etter uke. Denne figuren viser at det ikke er nevneverdige forskjeller i feilen fra uke til uke. Det vil derfor i denne analysen bli sett bort fra at det er forskjellige tidspunkter for undersøkelsene. En mer utfyllende analyse av dette problemet kommer i avsnitt 3.5.

Figur 3.4.1. Sysselsatt i AKU, men ikke i registeret i prosent av sysselsatt i AKU. uke



3.5. Analyse av mulige feilkilder i AKU og arbeidsmarkedsregisteret

Feilkildene i både AKU og arbeidsmarkedsregister kan bli analysert ved hjelp av logistisk regresjon. I modellen som blir satt opp her, er det valgt å bruke 'indirekte/direkte intervju', uke og registerkilde som bakgrunnsvariablene. Disse variablene er antatt å kunne best vise variasjonen i kvaliteten av statistikken. Responsvariabelen er samsvar i sysselsetningsstatus eller ikke. En slik respons variabel viser totalfeil mellom AKU og arbeidsmarkedsregisteret, men ikke hvilken retning feilen har. Det vil si om det går fra sysselsatt i AKU til ikke sysselsatt i arbeidsmarkedsregisteret eller omvent.

Responsvariabel

$Y=1$ hvis det er samsvar i sysselsetningsstatus i AKU og arbeidsmarkedsregister, ellers $Y=0$

Bakgrunnsvariablene (x)

Registerkildene i arbeidsmarkedsregisteret (regk):		
Kode	Forklaring	Antall
0	Ikke registrert i kilderegistrene	5 354
1	Koplet lønstakerforhold (AA mot LTO)	12 850
2	Ukoplet arbeidstakerforhold (AA uten LTO)	215
3	Ukoplet lønstakerforhold (kun LTO)	1 285
4	Selvangivelsesregister med inntekt fra primærnæringen	394
5	Selvangivelsesregister med inntekt fra andre næringer	574
6	Verneplikt og siviltjenesteregister	61
7	Sofa søkerregisteret	383

Indirekte/direkte intervju i AKU (akui)		
Kode	Forklaring	Antall
1	Direkte intervju	18 219
5	Indirekte intervju	2 927

Uke for intervju i AKU, med omtrent 1 600 intervju per uke.

Uke 44 ble valgt som referanseuke, dette var tellingsuka. For registerkilde var referansegruppa valgt til å være koplet lønnskaterforhold (AA mot LTO), fordi denne kilden er den største og beste kilden. For intervju var direkte intervju valgt som referansegruppe side direkte intervju er ansett for å være det mest korrekte.

Modell

odds(x)=

$$\frac{\pi(y = 1 | x)}{1 - \pi(y = 1 | x)} = \exp\left\{\alpha + \beta_{akui}x_{akui} + \beta_{regk0}x_{regk0} \cdots + \beta_{regk7}x_{regk7} + \beta_{uke40}x_{uke40} + \cdots + \beta_{uke52}x_{uke52}\right\}$$

x er en dummyvariabel som enten kan ha verdien 0 eller 1. For eksempel hvis registerkilde er 2 da er $x_{regk2}=1$ og de resterende registerkildene $x_{regkj}=0$. Det vil være tilsvarende for de andre bakgrunnsvariablene.

I denne modellen vil parameterne β bety effekten av en av bakgrunnsvariablene, i den spesielle kategorien, i forhold til referansegruppen. For eksempel betyr β_{regk2} er effekten av kilde 2 i forhold til referansegruppa som er kilde 1. Tilsvarende vil det være for de andre bakgrunnsvariablene.

$\pi(Y = 1 | x)$ er sannsynligheten for samsvar mellom AKU og register gitt bakgrunnsvariablene x.

Parameterne β blir estimert ved hjelp av maksimum likelihood metoden.

For å sammenligne forholdene mellom grupper kan man se på oddsforholdet

$$\frac{odds(X_{regk0})}{odds(x_{regk1})} = \exp(\hat{\beta}_{regk0})$$

Denne kan tolkes som forholdet mellom registerkilde 0 og 1. Hvis oddsforholdet er 1 er det ingen forskjell mellom kildene. For å teste dette kan vi se på 95% konfidensintervallet til oddsforholdet som er gitt ved:

$$KI = \exp(\hat{\beta}_{regk0} \pm 1,96 \cdot \sqrt{\text{vâr}(\hat{\beta}_{regk0})})$$

Tabell 3.5.1. Analyse av maksimum likelihood estimatene

Parameter	DF	Estimat	Standard feil	Wald Kji-kvadrat	Pr > Kji-kv
	1	3,22	0,10	1124,43	0,00
Indirekte intervju	1	-0,39	0,06	38,24	0,00
Ikke registrert i kilderegistrene	1	-1,40	0,06	557,23	0,00
Ukoplek arbeidstakerforhold (AA uten LTO)	1	-0,57	0,27	4,44	0,04
Ukoplek lønnstakerforhold (kun LTO)	1	-2,41	0,07	1039,10	0,00
Selvangivelsesregister med inntekt fra primærnæringen	1	-1,57	0,14	121,62	0,00
Selvangivelsesregister med inntekt fra andre næringer	1	-0,94	0,15	40,77	0,00
Verneplikt og siviltjenesteregister	1	-1,13	0,38	8,70	0,00
Sofa søkerregisteret	1	-2,01	0,13	244,73	0,00
Uke 40	1	0,07	0,12	0,30	0,58
Uke 41	1	-0,10	0,12	0,67	0,41
Uke 42	1	0,03	0,12	0,06	0,81
Uke 43	1	0,02	0,12	0,04	0,85
Uke 45	1	-0,05	0,12	0,19	0,66
Uke 46	1	0,14	0,13	1,14	0,29
Uke 47	1	-0,05	0,12	0,18	0,68
Uke 48	1	0,06	0,13	0,22	0,64
Uke 49	1	-0,12	0,12	0,91	0,34
Uke 50	1	0,14	0,13	1,24	0,26
Uke 51	1	-0,16	0,12	1,72	0,19
Uke 52	1	0,02	0,13	0,04	0,85

Testing av modellen

Hosmer og Lemshow goodness-of-fit test:

Kji-kvadrat : 39

Frihetsgrader : 8

Pr.>kji-kvadrat: <0,0001

Oppsummering av resultatene:

- Modellen er godt tilpasset dataene. Dette betyr at bakgrunnsvariablene som er valgt beskriver godt samsvaret (eller det manglende samsvaret) mellom AKU og arbeidsmarkedsregisteret.
- Når det er utført indirekte intervju synker samsvaret mellom AKU og registeret i klassifiseringen av sysselsatt med 30 prosent i forhold til direkte intervju. Da er det tatt hensyn til registerkilder og uke for intervju.
- Koplek lønnstakerforhold (kilde 1) er den beste kilden for registersyssetning. Alle andre registerkilder har klart dårligere kvalitet.
- Hvilken uke som er referansetidspunktet i AKU har ingen betydning for sammenligningen med arbeidsmarkedsregisteret, som er fra uke 44. Ved å ta bort variabelen 'uke' fra modellen, ble modellen bedre tilpasset data. Grunnen til at det at det ikke blir forskjell er at de fleste har et arbeidsforhold som varer i hele kvartalet.

Tabell 3.5.2. Estimer for oddsforholdet

	Punkt estimat	95% Wald konfidens grense
Direkte intervju versus indirekte intervju	0,68	0,60 0,77
Ikke registrert i kilderegistrene versus koplet lønnstakerforhold (AA mot LTO)	0,25	0,22 0,28
Ukoplet arbeidstakerforhold (AA uten LTO) versus koplet lønnstakerforhold (AA mot LTO)	0,56	0,33 0,96
Ukoplet lønnstakerforhold (kun LTO) versus koplet lønnstakerforhold (AA mot LTO)	0,09	0,08 0,10
Selvangivelsesregister med inntekt fra primærnæringen versus koplet lønnstakerforhold (AA mot LTO)	0,21	0,16 0,28
Selvangivelsesregister med inntekt fra andre næringer versus koplet lønnstakerforhold (AA mot LTO)	0,39	0,29 0,52
Verneplikt og siviltjenesteregister versus koplet lønnstakerforhold (AA mot LTO)	0,32	0,15 0,68
Sofa søkerregisteret versus koplet lønnstakerforhold (AA mot LTO)	0,13	0,10 0,17
Uke 40 versus 44	1,07	0,84 1,37
Uke 41 versus 44	0,91	0,71 1,15
Uke 42 versus 44	1,03	0,81 1,31
Uke 43 versus 44	1,02	0,80 1,31
Uke 45 versus 44	0,95	0,74 1,21
Uke 46 versus 44	1,15	0,89 1,47
Uke 47 versus 44	0,95	0,74 1,21
Uke 48 versus 44	1,06	0,83 1,36
Uke 49 versus 44	0,89	0,70 1,13
Uke 50 versus 44	1,15	0,90 1,49
Uke 51 versus 44	0,85	0,67 1,08
Uke 52 versus 44	1,02	0,80 1,32

3.6. Konsekvenser av bruk av indirekte intervju i AKU

Omtrent 14 prosent av intervjuene i AKU blir gjort indirekte. Blant de som er intervjuet indirekte er det en overhyppighet av feilklassifiserings som ikke sysselsatt i AKU i forhold til de som er intervjuet direkte (tabell 3.6.1). Dette utgjør omtrent 120 personer flere klassifisert som sysselsatt. Skjevheten mellom AKU og AKU med registervariabler er på 0,6 prosentpoeng. Hvis vi bare ser på de som har hatt et direkte intervju i AKU er skjevheten på bare 0,1 prosentpoeng. For de som har hatt indirekte intervju er skjevheten på hele 3,7 prosentpoeng. Overhyppighet av ikke sysselsatte i AKU skyldes trolig ikke feil i register fordi over halvparten i denne gruppen er registrert i både AA-registeret og i LTO med et arbeidsforhold (tabell 3.6.3). Denne andelen er større for de med indirekte intervju som er sysselsatt ifølge registeret og ikke AKU. Når noen er både registrert i AA-register og i LTO kan man med stor sannsynlighet anta at de er sysselsatt. Blant de som var registrert med både AA og LTO forhold og direkte intervju var det 3 prosent som ikke var sysselsatt ifølge AKU. For de i den samme gruppen, men med indirekte intervju var det hele 9 prosent som ikke var sysselsatt ifølge AKU. Denne målefeilen i AKU er større enn utvalgsfeilen og gjør at andelen sysselsatte i 4 kvartal 2001 burde være omtrent 71,4 og ikke 70,8. Men det kan være andre feil i AKU, slik som frafall blant innvandrere, som gjør at sysselsettingen burde vært lavere.

Tabell 3.6.1 Sysselsatt i register og AKU etter indirekte/direkte intervju. Vektet

	I alt	Sysselsatt i register og i AKU	Sysselsatt i register og ikke i AKU	Ikke sysselsatt men i AKU	Ikke sysselsatt og heller ikke i AKU	I alt	Sysselsatt i register og i AKU	Sysselsatt i register og ikke i AKU	Ikke sysselsatt men i AKU	Ikke sysselsatt og heller ikke i AKU
I alt	3 210 177	2 129 617	164 344	144 952	771 264	100,0	66,34	5,12	4,52	24,03
Direkte intervju	2 768 318	1 880 643	125 975	123 097	638 603	100,0	67,93	4,55	4,45	23,07
Indirekte intervju	441 859	248 974	38 369	21 855	132 661	100,0	56,35	8,68	4,95	30,02

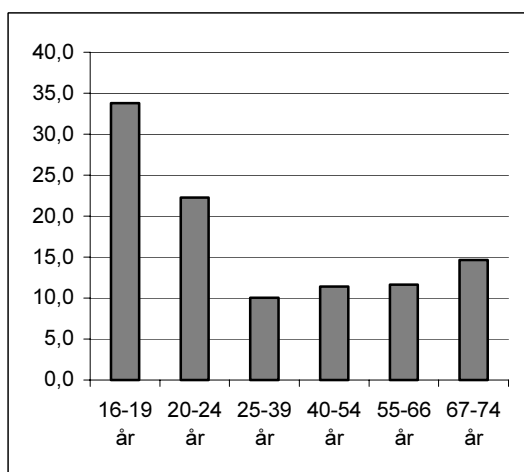
Tabell 3.6.2. Sysselsatt i register og AKU etter indirekte/direkte intervju og registerkilde

Antall	Direkte intervju					Indirekte intervju				
	I alt	Sysselsatt i register og i AKU	Sysselsatt i register og ikke i AKU	Ikke sysselsatt i register, men i AKU	Ikke sysselsatt i register og heller ikke i AKU	I alt	Sysselsatt i register og i AKU	Sysselsatt i register og ikke i AKU	Ikke sysselsatt i register, men i AKU	Ikke sysselsatt i register og heller ikke i AKU
I alt	18199	12612	827	749	4011	2913	1686	254	135	838
Ikke registrert i kilderegistrene	4422	0	0	668	3754	928	0	0	125	803
Koplet lønntakerforhold (AA mot LTO)	11275	10892	383	0	0	1575	1436	139	0	0
Ukoplet arbeidstakerforhold (AA uten LTO)	190	179	11	0	0	25	21	4	0	0
Ukoplet lønntakerforhold (kun LTO)	1070	752	318	0	0	215	119	96	0	0
Selvangivelsesregister med inntekt fra primærnæringen	353	295	58	0	0	41	33	8	0	0
Selvangivelsesregister med inntekt fra andre næringer	519	468	51	0	0	55	50	5	0	0
Verneplikt og siviltjenesteregister	32	26	6	0	0	29	27	2	0	0
Sofa søkerregisteret	338	0	0	81	257	45	0	0	10	35
Prosent										
I alt	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
Ikke registrert i kilderegistrene	24,3	0,0	0,0	89,2	93,6	31,9	0,0	0,0	92,6	95,8
Koplet lønntakerforhold (AA mot LTO)	62,0	86,4	46,3	0,0	0,0	54,1	85,2	54,7	0,0	0,0
Ukoplet arbeidstakerforhold (AA uten LTO)	1,0	1,4	1,3	0,0	0,0	0,9	1,2	1,6	0,0	0,0
Ukoplet lønntakerforhold (kun LTO)	5,9	6,0	38,5	0,0	0,0	7,4	7,1	37,8	0,0	0,0
Selvangivelsesregister med inntekt fra primærnæringen	1,9	2,3	7,0	0,0	0,0	1,4	2,0	3,1	0,0	0,0
Selvangivelsesregister med inntekt fra andre næringer	2,9	3,7	6,2	0,0	0,0	1,9	3,0	2,0	0,0	0,0
Verneplikt og siviltjenesteregister	0,2	0,2	0,7	0,0	0,0	1,0	1,6	0,8	0,0	0,0
Sofa søkerregisteret	1,9	0,0	0,0	10,8	6,4	1,5	0,0	0,0	7,4	4,2

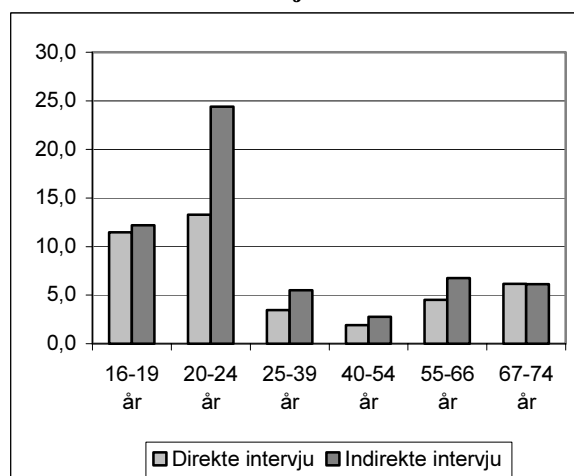
Konsekvenser av indirekte intervju for sysselsatte etter alder

Indirekte intervju forekommer hyppigst blant personer under 25 år, se figur 1. Omtrent 35 prosent av de mellom 16 og 19 år har intervjuet blitt gjort indirekte, mens det for de mellom 20-24 år var 22 prosent. Forskjellen mellom disse to gruppene er at de mellom 16 og 19 år i hovedsak bor hjemme og går på videregående skole, mens de mellom 20-25 år i hovedsak studerer og bor borte fra hjemmet. Hvis vi ser på andelen som er sysselsatt ifølge registeret, men ikke i AKU etter alder er det forholdsvis likt mellom de som har hatt indirekte og direkte intervju, bortsett fra de i aldersgruppen 20-24 år (figur 2). Dette tyder på at antall sysselsatte i alder 20-24 år i AKU er underrepresentert. Hvis vi antar at det er direkte intervju som har det korrekte forholdet mellom AKU og registeret vil andelen sysselsatt i alderen 20-24 år være 69,0 prosent og ikke 66,4, som er det tallet som ble publisert.

Figur 3.6.1. Andelen med indirekte intervju etter alder. Prosent



Figur 3.6.2. Andelen som er sysselsatt i registeret og ikke i AKU etter alder. Indirekte/direkte intervju. Prosent



Tabell 3.6.3. Andelen som er sysselsatt i registeret og ikke i AKU etter alder. Indirekte/direkte intervju. Antall og prosent

	Direkte intervju					Indirekte intervju				
	I alt	Sysselsatt i register og i AKU	Sysselsatt i register og ikke i AKU	Ikke sysselsatt i register, men i AKU	Ikke sysselsatt i register og heller ikke i AKU	I alt	Sysselsatt i register og i AKU	Sysselsatt i register og ikke i AKU	Ikke sysselsatt i register, men i AKU	Ikke sysselsatt i register og heller ikke i AKU
I alt	18199	12612	827	749	4011	2913	1686	254	135	838
16-19 år	1011	361	116	102	432	516	181	63	47	225
20-24 år	1331	875	177	65	214	381	202	93	17	69
25-39 år	5700	4715	197	244	544	637	481	35	32	89
40-54 år	5577	4677	107	196	597	719	571	20	25	103
55-66 år	3154	1893	142	106	1013	415	243	28	12	132
67-74 år	1426	91	88	36	1211	245	8	15	2	220
I alt	100,0	69,3	4,5	4,1	22,0	100,0	57,9	8,7	4,6	28,8
16-19 år	100,0	35,7	11,5	10,1	42,7	100,0	35,1	12,2	9,1	43,6
20-24 år	100,0	65,7	13,3	4,9	16,1	100,0	53,0	24,4	4,5	18,1
25-39 år	100,0	82,7	3,5	4,3	9,5	100,0	75,5	5,5	5,0	14,0
40-54 år	100,0	83,9	1,9	3,5	10,7	100,0	79,4	2,8	3,5	14,3
55-66 år	100,0	60,0	4,5	3,4	32,1	100,0	58,6	6,7	2,9	31,8
67-74 år	100,0	6,4	6,2	2,5	84,9	100,0	3,3	6,1	0,8	89,8

4. Målefeil og utvalgsfeil

Systematisk målefeil (skjevhet)

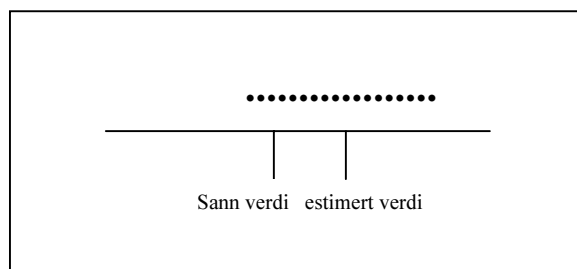
Det vil i enhver undersøkelse forekomme målefeil. Systematiske målefeil er den feilen som oppstår når det i undersøkelsen stadig blir målt enten for høy eller for lav verdi i forhold til sannheten, se figur 4.1. (Den prikkene linjen representere datapunkter) . Det vil si at det er en konstant skjevhet i utvalget eller registeret. Det er mulig å måle den systematiske målefeilen hvis en har to undersøkelser som er sammenlignbare, i dette notatet er det arbeidsmarkedsregisteret og utvalgsundersøkelsen AKU som skal sammenlignes.

Den systematiske målefeilen blir beregnet ved:

$$(4.1) \hat{\beta} = \hat{p}_{telling} - \hat{p}_{urvalg}$$

der \hat{p} er den estimerte andelen med en egenskap i henholdsvis arbeidsmarkedsregisteret og arbeidskraftsundersøkelsen AKU. Den estimerte systematiske målefeilen er et uttrykk for skjevheten i tellingen, hvis vi antar at arbeidskraftsundersøkelsen er sannheten. Dette er en antakelse som ikke er holdbar, siden det vil være målefeil i begge kilder. Men siden det ikke finnes noen bedre kilde til å beregne målefeilen vil AKU bli benyttet. Den estimerte systematiske målefeilen kan også bare bli sett på som forskjellen mellom resultatene i arbeidsmarkedsregisteret og AKU. Hovedproblemet hvis det eksisterer skjevheter (systematisk målefeil) er at det ikke hjelper med å øke utvalgsstørrelsen eller å benytte fulltelling. Skjevheten vil uansett være like stor.

Figur 4.1 Systematisk målefeil- en visuell forklaring



Tilfeldig målefeil

Tilfeldig målefeil er den feilen som oppstår i undersøkelser når målingene bommer tilfeldig rundt den sanne verdien gitt at det ikke er en systematisk målefeil, se figur 4.2. (Den prikkene linjen representere datapunkter) Det vil si noen ganger under den sanne verdien, mens andre ganger over den sanne verdien. I gjennomsnitt blir det bommet like mange ganger over som under den sanne verdien. Dette er i motsetning til den systematiske målefeilen som i hovedsak enten bommer over eller under sannheten. Den tilfeldige målefeilen vil bli liten hvis vi øker utvalgsstørrelsen eller bruker fulltelling. Hvis målingen blir gjort mange ganger, vil gjennomsnittet konvergere mot den sanne verdien.

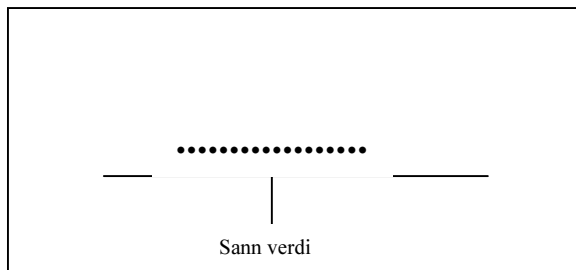
Den tilfeldige målefeilen blir her beregnet ved:

$$(4.2) \hat{\tau}^2 = \hat{p}_{01} + \hat{p}_{10} - (\hat{p}_{01} - \hat{p}_{10})^2$$

der :

- \hat{p}_{01} er andelen som ikke har egenskap i arbeidsmarkedsregisteret, men har den i AKU
- \hat{p}_{10} er andelen som har egenskapen i arbeidsmarkedsregisteret, men ikke i AKU.

Figur 4.2. Tilfeldig målefeil- en visuell forklaring



Relativ målefeil

Både den estimerte tilfeldige og systematiske målefeilen kan ikke si oss mye uten at de er sett i forhold til selve estimatet, derfor vil det i analysen bli sett på den relative målefeilen. De relative målefeilene forteller hvor stor feilen er i prosent av størrelsen vi har estimert. Når den relative feilen er på over 20 prosent er den ofte ansett som stor.

Den relative totale målefeilen er forholdet mellom den totale målefeilen og størrelsen på estimatet. Den relative totale målefeilen er gitt ved:

$$(4.3) \quad R_{Målefeil} = \frac{100 \times \sqrt{\frac{\hat{\epsilon}^2}{n} + \hat{\beta}^2}}{\hat{p}_{telling}}$$

n er antall personer i populasjonen.

Den relative systematiske målefeilen (skjevheten) er gitt ved:

$$(4.4) \quad R_{Systematisk} = \frac{100 \times \hat{\beta}}{\hat{p}_{telling}}$$

Dette er et uttrykk for hvor stor skjevheten er i forhold til estimatet i prosent.

Den relative tilfeldige målefeilen er gitt ved:

$$(4.5) \quad R_{tilfeldig} = \frac{100 \times \frac{\hat{\epsilon}}{\sqrt{n}}}{\hat{p}_{telling}}$$

n er antall personer i populasjonen.

For nærmere utredning av formlene, se upublisert notat av Solheim et al.

Utvalgfeil (for AKU utvalget)

Utvalgsfeilen blir beregnet på følgende måte:

$$(4.8) \quad \hat{Var}(\hat{p}) \approx \text{var}\left(\frac{X}{n}\right) = \frac{1}{n^2} \cdot \text{var}(X) = \frac{n \cdot \hat{p} \cdot (1 - \hat{p})}{n^2} = \frac{\hat{p} \cdot (1 - \hat{p})}{n} = \frac{\hat{\sigma}^2}{n}$$

I formelen (4.8) er det antatt at AKU er et enkelt tilfeldig utvalg. AKU er imidlertid et familieutvalg det vil si et klyngeutvalg og ikke et enkelt tilfeldig utvalg. I tillegg er fylker med få innbyggere overrepresentert i forhold til folketallet og vektene er kalibrert i forhold til registervariabler som er kjent for hele populasjonen. Ifølge et internt notat av Solheim og Zhang blir standardavviket (roten av

formelen 4.8) redusert med 50 prosent hvis vi i variansberegningene tar hensyn til kalibreringen. Imidlertid fører frafallet i AKU til skjevheter i andelen sysselsatte på mellom 1 og 2 standardavvik. Ut fra dette ble det valgt å beregne utvalgtfeilen slik som i formel (4.8)

95 prosent konfidensintervallet er gitt ved:

$$(4.9) \left(\hat{p} - 2 \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}}, \hat{p} + 2 \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}} \right)$$

5. Analyse av kvaliteten

5.1. Sysselsatt

Metoden som brukes for å finne sysselsetningsstatus til personene bygger på at totalt antall sysselsatte fra registeret skal være likt totalt antall sysselsatte i AKU. Det er derfor ingen forskjell for antall sysselsatte i arbeidsmarkedsregisteret og AKU. Metoden låser ikke antall sysselsatte menn og kvinner til samme nivå som AKU. Det kan derfor være en forskjell mellom arbeidsmarkedsregisteret og AKU for antall sysselsatte menn og kvinner. Forskyvingen mellom kjønnene er ikke stor, bare omtrent 0,2 prosentpoeng.

Hvis vi bruker AKU utvalget, med registervariabler, blir bildet noe annerledes. Da blir sysselsetningen høyere for både menn og kvinner. AKU utvalget ser dermed ikke ut til å være representativt i forhold til registersysselsetning. Grunnen til dette skyldes antakelig målefeil og frafall i AKU, se avsnitt 3.3 og 3.6. AKU med registervariabler ligger innenfor 95 prosent konfidensintervall til AKU med intervjuvariabler, se tabell 5.1.1 og 5.1.2.

Tabell 5.1.1. Personer 16-74 år, etter sysselsetning. Arbeidsmarkedsregister og AKU. Prosent

	I alt	Menn	Kvinner
Arbeidsmarkedsregister 29. okt. - 4nov. 2001	70,8	74,7	66,9
AKU 4.kvartal 2001	70,8	74,5	67,1
AKU 4. kvartal 2001 med registervariabler	71,4	75,3	67,6

I analysen videre er det AKU utvalget som vil bli benyttet. Det vil i hovedsak bli sett på forskjellen når det er benyttet registervariabler og variabler fra selve intervjuet. Dette vil være tilnærmet det samme som å se på forskjellen mellom AKU og hele arbeidsmarkedsregisteret. I enkelte tilfeller vil målefeil (AKU), utvalgsfeil, frafall eller vekting gjøre at det blir forskjell mellom arbeidsmarkedsregisteret og AKU med registervariabler.

For sysselsatt etter kjønn er skjevheten, det vil si den systematiske målefeilen, større enn utvalgsfeilen, se tabell 5.1.2. Likevel er den relative skjevheten i forhold til estimatet på bare 1 prosent og mindre. Den tilfeldige målefeilen er liten og den er mindre enn utvalgsfeilen. Dette vil si at kvaliteten for totaltall for sysselsetning etter kjønn er akseptabel.

Tabell 5.1.2. Utvalgsfeil og målefeil for sysselsetning etter kjønn. AKU med registervariabler

	Antall personer i utvalget	Andelen,		Utvalgsfeil		Systematisk målefeil		Tilfeldig målefeil	
		\hat{p}	$\hat{\sigma} / \sqrt{n}$	Relativ utvalgfeil, prosent	$\hat{\beta}$	Relativ målefeil, prosent	$\hat{\tau} / \sqrt{n}$	Relativ målefeil, prosent	
I alt	21116	0,714	0,003	0,4	0,006	0,8	0,002	0,0	
Menn	10633	0,753	0,004	0,6	0,008	1,0	0,003	0,0	
Kvinner	10483	0,676	0,005	0,7	0,005	0,7	0,003	0,0	

5.1.1. Kvaliteten av kildedataene

Arbeidsmarkedsregisteret er satt sammen av opplysninger fra mange forskjellige kilder. Kvaliteten mellom de forskjellige kildene varierer, se tabell 5.1.3. Den største og beste kilden for sysselsetingsdata er koplet lønnstakerforhold mellom arbeidstakerregisteret (AA) og lønns- og trekkoppgaverregisteret (LTO). Det er bare fire prosent av de som er klassifisert som sysselsatt i denne gruppen som i AKU ikke er sysselsatt. Dette vil si at skjvheten (den relative systematiske målefeilen) er på 4 prosent. Den tilfeldige målefeilen er også liten, den relative tilfeldige målefeilen er på 0,2 prosent.

Den nest største kilden til sysselsetningen er de som bare har opplysninger i LTO og ikke i AA. Av de som er klassifisert som sysselsatt ifølge denne kilden er det 31,6 prosent som ikke er sysselsatt ifølge AKU. Altså er skjvheten (den relative systematiske målefeilen) svært stor.

De som bare var registrert i AA og ikke LTO er det bare 6,5 prosent som er klassifisert som sysselsatt ifølge registeret, men som ikke er det ifølge AKU. De har en liten systematisk feil, men det er få personer som bare er registrert i AA.

De som har selvangivelsesregisteret som eneste kilde til sysselsetning er det ganske stor forskjell mellom register og AKU, spesielt de med inntekt fra primærnæring. For denne registerkilden er det 18 prosent som er klassifisert som sysselsatt i registeret og som ikke er det i AKU. For de med inntekt fra andre næringer er det 10 prosent som er sysselsatt ifølge denne kilden, men som ikke er det ifølge AKU.

Blant de som var registrert i Sofa-søkerregister og derfor klassifisert som arbeidsledig, var det 24 prosent som var sysselsatt ifølge AKU.

Blant de som ikke hadde noen registeropplysninger og ikke ble klassifisert som sysselsatt, var det 15 prosent som i AKU oppga at de var sysselsatt.

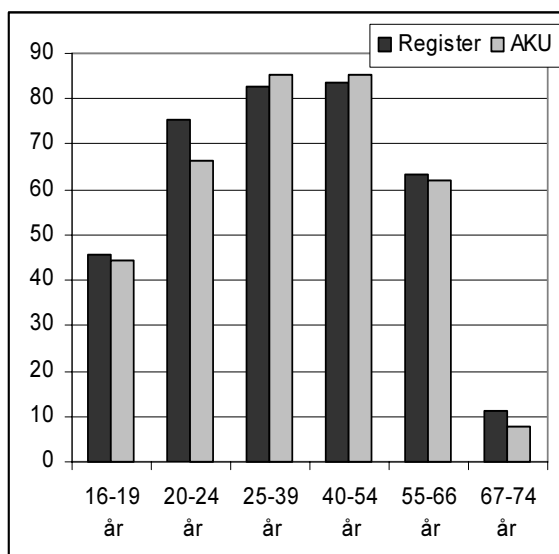
Tabell 5.1.3. Systematisk, tilfeldig og relativ målefeil for sysselsetning etter arbeidsmarkedskilde

Arbeidsmarkedskilde	Antall personer i utvalget	Andelen, \hat{p}	Systematisk målefeil, $\hat{\beta}$	Relativ systematisk målefeil, prosent	Tilfeldig målefeil, $\hat{\tau} / \sqrt{n}$	Relativ tilfeldig målefeil, prosent
Ikke registrert i kilderegistrene						
Koplet lønnstakerforhold (AA mot LTO)	5354	0	-0,152	-	0,005	-
Ukoplet arbeidstakerforhold (AA uten LTO)	12850	1	0,040	4,0	0,002	0,2
Ukoplet lønnstakerforhold (kun LTO)	215	1	0,065	6,5	0,017	1,7
Selvangivelsesregister med inntekt fra primærnæringen	1285	1	0,316	31,6	0,013	1,3
Selvangivelsesregister med inntekt fra andre næringer	394	1	0,178	17,8	0,019	1,9
Verneplikt og siviltjenesteregister	574	1	0,101	10,1	0,013	1,3
Sofa søkerregisteret	61	1	0,122	12,2	0,042	4,2
	383	0	-0,235	-	0,022	-

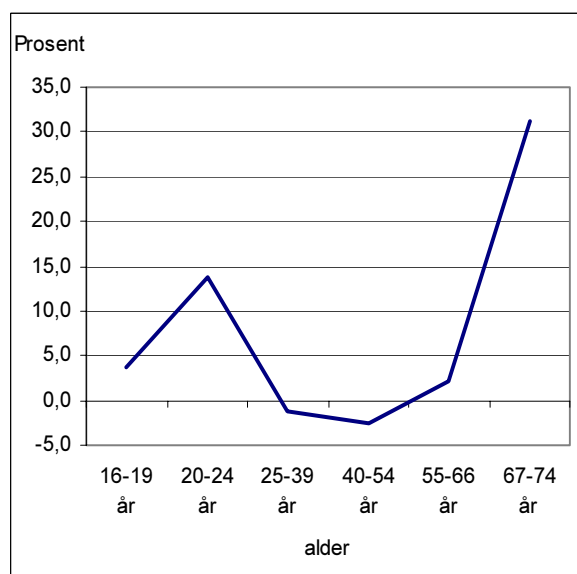
5.1.2. Andelen sysselsatte etter alder

For de som er i alderen 25-66 år er det svært liten forskjell i andelen sysselsatte mellom AKU og arbeidsmarkedsregisteret. For de i aldersgruppen 20-24 år er det 10 prosentpoeng høyere sysselsetning ifølge arbeidsmarkedsregisteret enn i forhold til AKU, men AKU ligger omtrent 2 prosentpoeng for lavt på grunn av målefeil i forbindelse med indirekte intervju, se avsnitt 3.6. Den reelle forskjellen er da på omtrent 8 prosentpoeng. For de i aldersgruppen 67-74 år er andelen sysselsatte 3 prosentpoeng høyere i arbeidsmarkedsregisteret enn i AKU. Andelen sysselsatte for begge disse gruppene ligger utenfor 95 prosent konfidensintervall til AKU. Grunnen til denne forskjellen er at det i disse aldersgruppene er det mange som ikke har faste jobber og det er derfor vanskelig å klassifisere om de arbeidet eller ikke på tellingstidspunktet.

Figur 5.1.1. Andelen sysselsatte etter alder. Arbeidsmarkedsregisteret og AKU



Figur 5.1.2. Relativ skjevhet etter alder. AKU utvalget



Tabell 5.1.4. Andelen sysselsatte etter alder. Arbeidsmarkedsregisteret og AKU. 2001

	Arbeidsmarkeds-	AKU 4. kvartal	AKU 4.kvartal 2001			
	register	2001 med	Estimert andel	Standardavvik	Nedre grense	Øvre grense
	andel	Estimert andel	Estimert andel		95 %	95 %
					konfidensintervall	
16-19 år	0,46	0,46	0,45	0,01	0,42	0,47
20-24 år	0,76	0,77	0,66	0,01	0,64	0,69
25-39 år	0,83	0,84	0,85	0,00	0,84	0,86
40-54 år	0,84	0,83	0,85	0,00	0,84	0,86
55-66 år	0,63	0,64	0,62	0,01	0,61	0,64
67-74 år	0,11	0,11	0,08	0,01	0,07	0,09

I tillegg hvis vi ser på den systematiske og tilfeldige målefeilen (tabell 5.1.5 og figur 5.1.1) ser vi at den er liten for de i aldersgruppen 25-66 år, mens den er stor for de i alderen 20-24 år og 67-74 år. For de i alderen 20-24 år er 14 prosent flere som er klassifisert som sysselsatt i registeret enn ifølge AKU. Tallene fra AKU ligger imidlertid noe for lavt på grunn av indirekte intervju, se avsnitt 3.6. Det er 31

prosent flere av de i alderen 67-74 år som er klassifisert som sysselsatt i arbeidsmarkedsregisteret, enn det er i AKU. Det er i denne 8 prosent som er sysselsatt i denne aldersgruppen ifølge AKU, mens det er 11 prosent ifølge arbeidsmarkedsregisteret. Grunnen til at den relative målefeilen blir stor er at det få sysselsatte i denne gruppen i forhold til størrelsen på skjevheten.

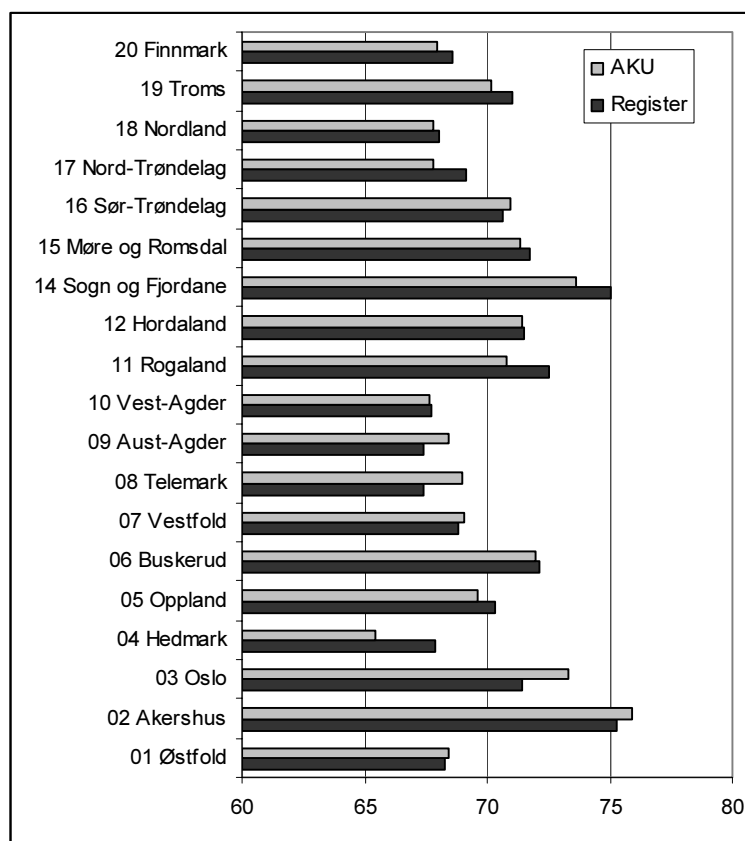
Tabell 5.1.5. Målefeil for sysselsetning etter alder arbeidsmarkedsregisteret. AKU-utvalget

	Antall personer i utvalget	Andelen	Systematisk målefeil		Tilfeldig målefeil	
		\hat{p}	$\hat{\beta}$	Relativ målefeil, prosent	$\hat{\tau} / \sqrt{n}$	Relativ målefeil, prosent
16-19 år	1527	0,46	0,02	3,7	0,012	2,6
20-24 år	1712	0,77	0,11	13,7	0,011	1,4
25-39 år	6337	0,84	-0,01	-1,2	0,004	0,4
40-54 år	6296	0,83	-0,02	-2,4	0,003	0,4
55-66 år	3569	0,64	0,01	2,1	0,005	0,8
67-74 år	1675	0,11	0,04	31,2	0,007	5,9

5.1.3. Andelen sysselsatt etter fylke

Andelen sysselsatte per fylke er det laveste geografiske nivået som det blir gitt tall for i AKU. Konfidensintervallet blir større enn for landet, fordi det er færre personer i utvalget når det er fordelt på fylker. Bredden på 95 prosent konfidensintervallet ligger på mellom 4 og 8 prosentpoeng for hvert fylke. Andelen sysselsatte per fylke ifølge arbeidsmarkedsregisteret ligger innenfor konfidensintervallet til AKU. I enkelte fylker ligger tallet fra arbeidsmarkedsregisteret på grensen til AKUs konfidensintervall.

Figur 5.1.3. Andelen sysselsatte etter fylke. Arbeidsmarkedsregisteret og AKU



Tabell 5.1.6. Andelen sysselsatte etter fylke

	AKU 4.kvartal 2001					
	Arbeids- markeds- register 29.okt -4. nov	AKU 4. kvartal 2001 med register- variabler	Estimert andel	Standardavvik	Nedre grense konfidensintervall	Øvre grense konfidensintervall
	Andel	Estimert andel				
01 Østfold	0,68	0,70	0,68	0,01	0,66	0,71
02 Akershus	0,75	0,76	0,76	0,01	0,74	0,78
03 Oslo	0,71	0,73	0,73	0,01	0,71	0,75
04 Hedmark	0,68	0,66	0,65	0,02	0,62	0,68
05 Oppland	0,70	0,72	0,70	0,02	0,67	0,73
06 Buskerud	0,72	0,73	0,72	0,01	0,69	0,75
07 Vestfold	0,69	0,70	0,69	0,01	0,66	0,72
08 Telemark	0,67	0,68	0,69	0,02	0,66	0,72
09 Aust-Agder	0,67	0,68	0,68	0,02	0,65	0,72
10 Vest-Agder	0,68	0,66	0,68	0,02	0,64	0,71
11 Rogaland	0,73	0,72	0,71	0,01	0,69	0,73
12 Hordaland	0,72	0,71	0,72	0,01	0,69	0,74
14 Sogn og Fjordane	0,75	0,76	0,74	0,02	0,70	0,77
15 Møre og Romsdal	0,72	0,74	0,71	0,01	0,69	0,74
16 Sør-Trøndelag	0,71	0,71	0,71	0,01	0,68	0,74
17 Nord-Trøndelag	0,69	0,69	0,68	0,02	0,64	0,71
18 Nordland	0,68	0,69	0,68	0,01	0,65	0,71
19 Troms	0,71	0,73	0,70	0,02	0,67	0,74
20 Finnmark	0,69	0,70	0,68	0,02	0,64	0,72

Hvis vi ser på målefeil for arbeidsmarkedsregisteret (tabell 5.1.7) ser vi at den systematiske målefeilen er større i enkelte fylker enn andre. De med størst systematisk målefeil er fylkene Oppland, Sogn og fjordane, Møre og Romsdal, Nordland og Troms. Den relative skjevheten ligger på mellom 3 og 4 prosent. Grunnen til at disse fylkene får dårligere kvalitet er at de sannsynligvis har en høyere andel av de sysselsatte som har selvangivelses registeret som kilde eller at de sysselsatte bare er registrert i LTO- og ikke AA-registeret. Den tilfeldige målefeilen er det ikke stor forskjell på imellom fylkene. På kommunenivå vil kvaliteten på tallene avhenge av hvor mange som har kildene selvangivelsesregisteret og bare LTO i arbeidsmarkedsregisteret. I kommuner hvor det er en høy andel av personer som jobber i et annet land vil sysselsetningen antakelig bli for lav fordi disse personene ikke er registrert i noen norske registre som gjelder arbeidsmarked.

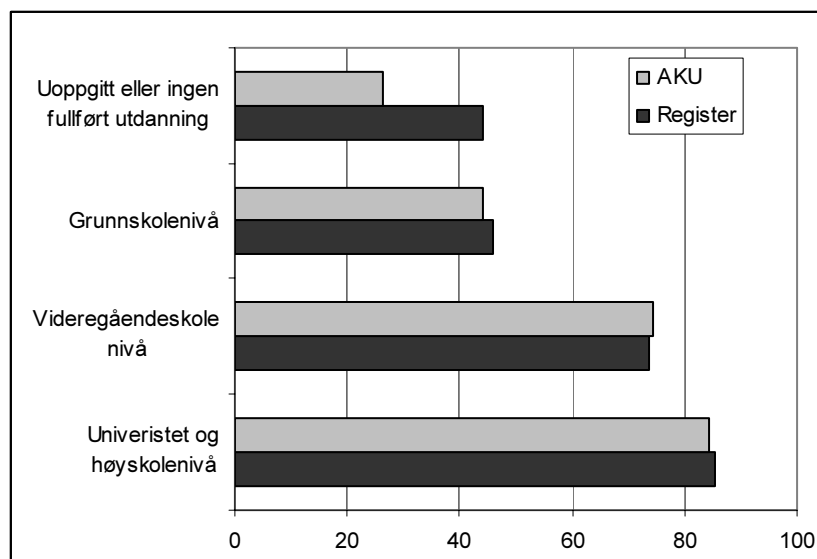
Tabell 5.1.7. Målefeil for sysselsetning etter fylke for arbeidsmarkedsregisteret. AKU-utvalget

	Antall personer i utvalget	Andelen		Systematisk målefeil		Tilfeldig målefeil	
		\hat{p}	$\hat{\beta}$	Relativ målefeil, prosent	$\hat{\tau} / \sqrt{n}$	Relativ målefeil, prosent	
01 Østfold	1 190	0,698	0,014	2,0	0,008	1,2	
02 Akershus	1 944	0,759	-0,000	-0,0	0,007	0,9	
03 Oslo	1 876	0,727	-0,006	-0,8	0,007	0,9	
04 Hedmark	959	0,661	0,007	1,0	0,010	1,5	
05 Oppland	877	0,723	0,027	3,7	0,011	1,5	
06 Buskerud	1 116	0,725	0,005	0,8	0,009	1,3	
07 Vestfold	1 022	0,699	0,008	1,2	0,010	1,4	
08 Telemark	789	0,683	-0,007	-1,0	0,011	1,6	
09 Aust-Agder	735	0,677	-0,007	-1,0	0,012	1,8	
10 Vest-Agder	729	0,663	-0,013	-2,0	0,012	1,8	
11 Rogaland	1 816	0,721	0,013	1,9	0,008	1,1	
12 Hordaland	1 799	0,712	-0,002	-0,3	0,007	0,9	
14 Sogn og Fjordane	715	0,760	0,024	3,2	0,011	1,5	
15 Møre og Romsdal	1 122	0,736	0,023	3,1	0,010	1,3	
16 Sør-Trøndelag	1 203	0,709	-0,000	-0,0	0,009	1,3	
17 Nord-Trøndelag	743	0,694	0,017	2,4	0,011	1,6	
18 Nordland	1 139	0,687	0,009	1,3	0,009	1,4	
19 Troms	691	0,728	0,026	3,6	0,013	1,8	
20 Finnmark	648	0,701	0,022	3,2	0,013	1,9	

5.1.4. Andelen sysselsatt etter utdanning

I AKU blir den høyeste utdanningen hentet fra Utdanningsregisteret før første intervju. Ved de senere intervjuene blir det spurt om personen har fullført en utdanning etter det forrige intervjuet. For de som har uoppgitt utdanning i utdanningsregisteret blir det i AKU spurt om utdanning ved første intervju og svaret blir kodet etter standarden for utdanning. Opplysningene om utdanning i AKU er derfor en blanding mellom intervjuopplysninger og registerdata.

Figur 5.1.4. Andelen sysselsatte etter utdanning. Arbeidsmarkedsregisteret og AKU



Det har utviklet seg en forskjell mellom utdanningsopplysninger i AKU og register (tabell 5.1.8). Det er hele 16 prosent som i AKU bare har grunnskole som høyeste utdanning, men som i utdanningsregisteret har videregående skolenivå som høyeste utdanning. Grunnen til dette skyldes antakelig at de som har fullført grunnkurs og videregående kurs 1 ikke har fortalt dette i intervjuet i AKU. Dette kan være fordi de ikke har ansett dette som fullført videregående utdanning og har misforstått spørsmålet. For de som bare har videregående skole i AKU, er det svært godt samsvar med utdanningsregisteret. For de som ifølge AKU har fullført universitet og høyskole utdanning er det 8 prosent som ifølge registeret ikke har gjort det. Grunnen til dette kan enten være at intervjuobjektet oppgir for høy utdanning eller at det er etterslep i statistikken, slik at disse opplysningene kommer senere inn i registeret.

I utdanningsregisteret var det på tellingstidspunktet 2,5 prosent av de mellom 16-74 år som ikke hadde noen utdanningsopplysninger registeret. Blant de som svarte i AKU var det 1,6 prosent som ikke hadde noen utdannings opplysninger i register. Disse blitt spurt om utdanningen sin og 2/3 av de har svarte, slik at det i AKU finnes opplysninger om utdanningen deres. Av de som har uoppgitt utdanning i registeret og som har fått kodet en utdanning i AKU, er det 29 prosent som bare har grunnskole, 26 prosent videregående skolenivå og 12 prosent universitet- eller høyskole nivå.

Tabell 5.1.8. Utdanningsnivå i AKU og i register

Register	AKU									
	I alt	Grunn- skole- nivå	Videre- gående- skolenivå	Univer- sitet- og høy- skolenivå	Uoppgitt eller ingen fullført utdann- ing	I alt	Grunn- skole- nivå	Videre- gående- skolenivå	Univer- sitet- og høy- skolenivå	Uoppgitt eller ingen fullført utdann- ing
I alt	21 116	4 203	11 280	5 503	130	100,0	19,9	53,4	26,1	0,6
Grunnskole	3 560	3 445	89	19	7	100,0	96,8	2,5	0,5	0,2
Videregåendesko- lenivå	12 115	660	10 994	446	15	100,0	5,4	90,7	3,7	0,1
Universitet- og høyskolenivå	5 111	1	111	4 998	1	100,0	0,0	2,2	97,8	0,0
Uoppgitt eller ingen fullført utdanning	330	97	86	40	107	100,0	29,4	26,1	12,1	32,4

Andelen sysselsatte etter utdanning ifølge arbeidsmarkedsregisteret ligger i grenseområdet til konfidensintervallet til AKU (tabell 5.1.9). Grunnen til dette skyldes forskjellen i utdanningsopplysningene i AKU og register. I AKU er det stort frafall blant de med uoppgitt utdanningsopplysninger i registeret, dette er i hovedsak innvandrere som ikke er sysselsatt, se avsnitt 3.3 om frafallet. I tillegg har de av disse som har svart, fått kodet en utdanning i AKU hvis de ikke hadde det fra før. Dette har ført til en forskyvning av andelen sysselsatte mellom utdanningsnivåene. Andelen sysselsatte på utdanningsnivå: grunnskole og universitet og høyskole er overestimert i registeret i forhold til AKU. Mens andelen sysselsatte på videregående skolenivå er underestimert i registeret i forhold til AKU. Den relative systematiske målefeilen er imidlertid svært liten. Grunnen til dette er at AKU med registervariabler ligger nærmere AKU enn arbeidsmarkedsregisteret.

Tabell 5.1.9. Sysselsatte etter utdanningsnivå. Arbeidsmarkedsregisteret og AKU

	Arbeidsmarkedsregisteret 29. okt-4. nov		AKU 4. kvartal 2001			
	Andel	Estimert andel	Estimert andel	Standardavvik	Nedre grense 95 % konfidensintervall	Øvre grense 95 % konfidensintervall
I alt	0,708	0,714	0,708	0,003	0,702	0,715
Grunnskole-nivå	0,458	0,448	0,440	0,008	0,425	0,456
Videregående skole-nivå	0,735	0,739	0,743	0,004	0,735	0,751
Universitet- og høyskolenivå	0,855	0,853	0,845	0,005	0,835	0,855
Uoppgitt eller ingen fullført utdanning	0,443	0,432	0,262	0,039	0,185	0,339

Tabell 5.1.10. Målefeil for sysselsatte etter utdanning for arbeidsmarkedsregister. AKU-utvalget

	Andel	Systematisk målefeil		Tilfeldig målefeil		
		Antall personer i utvalget	\hat{p}	$\hat{\beta}$	Relativ målefeil, prosent	\hat{t} / \sqrt{n}
Grunnskole-nivå	3560	0,448	0,001	0,3	0,005	1,2
Videregående skole-nivå	12115	0,739	0,008	1,1	0,003	0,4
Universitet- og høyskolenivå	5111	0,853	0,005	0,5	0,004	0,4
Uoppgitt eller ingen fullført utdanning	330	0,432	0,017	3,9	0,018	4,1

5.2. Yrkesstatus

Andelen selvstедige i forhold til antall sysselsatte er i arbeidsmarkedsregisteret satt lik andelen i AKU. Av de sysselsatte er det 6,6 prosent som er selvstедige næringsdrivende. Selv om andelen personer som er selvstедige er lik, er det ikke de samme personene som er selvstедige (tabell 5.2.1). 69 prosent av de som er selvstедige ifølge registeret, er det også ifølge AKU. 18 prosent er lønnstakere og 13 prosent er ikke sysselsatt. Grunnen til dette dårlige samsvaret på person nivå er antakelig fordi at mange blir klassifisert som selvstедige ut fra inntekt fra det foregående året og det trenger ikke å være det riktige for tellingsuka.

Tabell 5.2.1. Yrkesstatus i arbeidsmarkedsregister og AKU

Arbeidsmarkedsregister	AKU				AKU prosent			
	I alt	Ikke sysselsatt	Lønnstaker	Selvstедig	I alt	Ikke sysselsatt	Lønnstaker	Selvstедig
I alt	21 116	5 934	14 193	989	100,0	28,1	67,2	4,7
Ikke sysselsatt	5 737	4 853	686	198	100,0	84,6	12,0	3,5
Lønnstaker	14 411	959	13 332	120	100,0	6,7	92,5	0,8
Selvstедig	968	122	175	671	100,0	12,6	18,1	69,3

Den systematiske målefeilen for selvstедig er mindre enn for lønnstakere og ikke sysselsatte, men på grunn av andelen selvstедig er så liten blir den relative målefeilen større (tabell 5.2.2). De relative feilene er på mindre enn 5 prosent og det er akseptabelt.

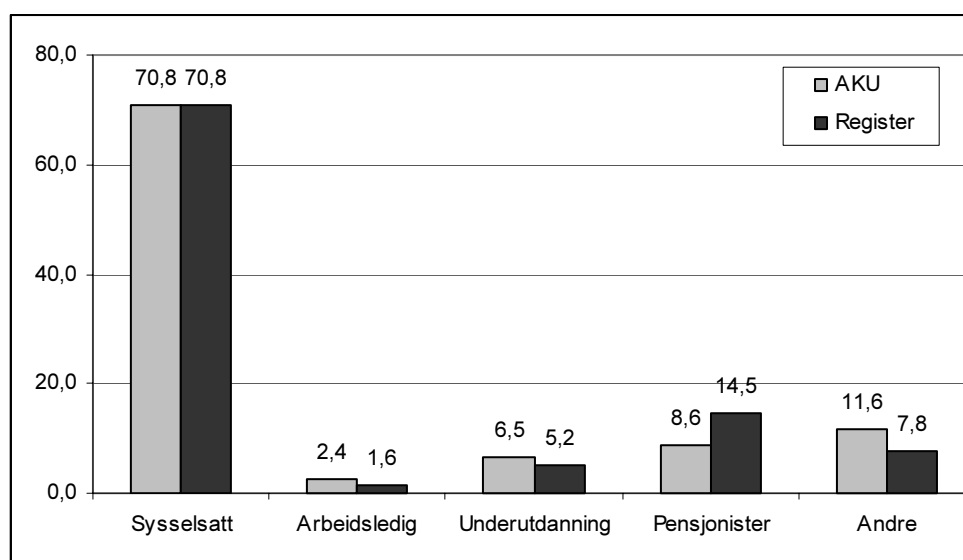
Tabell 5.2.2. Målefeil for yrkesstatus i arbeidsmarkedsregisteret. AKU-utvalget

	Antall personer i utvalget	Systematisk målefeil		Tilfeldig målefeil		
		Andel \hat{p}	$\hat{\beta}$	Relativ målefeil, prosent	$\hat{\tau} / \sqrt{n}$	Relativ målefeil, prosent
Ikke sysselsatt	5 737	0,286	-0,006	-2,1	0,002	0,7
Lønnstaker	14 411	0,670	0,008	1,2	0,002	0,3
Selvstendig	968	0,045	-0,002	-4,4	0,001	2,7

5.3. Aktivitetsstatus

Aktivitetsstatus er oppdelingen av personer etter sysselsatt, arbeidsledig og utenfor arbeidsstyrken. Hvor utenfor arbeidsstyrken i tillegg er oppdelt i under utdanning, pensjonist og andre.

Figur 5.3.1. Personer etter aktivitetsstatus. Arbeidsmarkedsregisteret og AKU



Tabell 5.3.1. Personer etter aktivitetsstatus i arbeidsmarkedsregisteret og AKU

	Tellingen 29. okt-4.nov	AKU 4. kvartal 2001 med register-variabler	AKU 4. kvartal 2001			
	Andel	Estimert andel	Estimert andel	Standardavvik	Nedre grense 95 % konfidens-intervall	Øvre grense 95 % konfidens-intervall
Sysselsatt	70,8	71,5	70,8	0,31	70,2	71,5
Arbeidsledig	1,6	1,8	2,4	0,10	2,2	2,6
Utenfor arbeidsstyrken:						
Under utdanning	5,2	5,4	6,5	0,17	6,2	6,9
Pensjonister	14,5	14,6	8,6	0,19	8,2	9,0
Andre	7,8	6,8	11,6	0,22	11,1	12,0

I arbeidsmarkedsregisteret er det et strengere kriterium for å bli klassifisert som arbeidsledig enn i AKU. I AKU er 2,4 prosent arbeidsledige av alle personer i alderen 16-74 år, mens det er 1,6 prosent i tellingen. Det er 6,5 prosent av befolkningen i alderen 16-74 år som er under utdanning i AKU mot 5,2

prosent i arbeidsmarkedsregisteret. I kategorien annet inngår gruppene arbeidsuføre og hjemmearbeidene. Det er stor forskjell mellom arbeidsmarkedsregisteret og AKU for denne kategorien, det kan tyde på at grensene for disse klassifiseringene pensjonister og andre er uklare.

For de i arbeidsmarkedsregisteret som er sysselsatt, er 93 prosent sysselsatt ifølge AKU, mens 3 prosent er under utdanning og resten er i de andre kategoriene (tabell 5.3.2.).

Det er bare 46 prosent av de som er arbeidsledige ifølge registeret som også er arbeidsledig ifølge AKU. Omtrent en fjerdedel av de som er arbeidsledig ifølge registeret er sysselsatt ifølge AKU. I tillegg er over en fjerdedel utenfor arbeidsstyrken.

For de som er under utdanning ifølge arbeidsmarkedsregisteret er 71 prosent det også i AKU, mens 17 prosent er sysselsatt.

Det er bare 5 prosent av de som er registrert som pensjonistene ifølge registeret som er sysselsatt ifølge AKU. Det er mange blant pensjonistene i registeret som har kommet i gruppen 'andre' i AKU. Dette kan skyldes en feilklassifisering enten i AKU eller i registeret.

Tabell 5.3.2. Aktivitetsstatus i arbeidsmarkedsregister og AKU

Register	Personer i utvalget	I alt	Sysselsatt	Arbeidsledig	AKU		
					Utenfor arbeidsstyrken		
					Under utdanning	Pensjonister	Andre
I alt	21 116	100,0	71,9	2,2	6,5	8,4	11,0
Sysselsatt	15 379	100,0	93,0	1,0	3,0	1,0	2,0
Arbeidsledig	338	100,0	24,0	45,9	1,5	0,9	27,8
Utenfor arbeidsstyrken:							
Under utdanning	1 126	100,0	16,7	8,5	70,6	0,0	4,2
Pensjonister	2 972	100,0	4,9	0,5	0,4	51,8	42,4
Andre	1 301	100,0	36,1	4,4	6,9	5,6	47,0

De relative systematiske målefeilene for arbeidsledige og alle kategoriene under arbeidsstyrken er svært stor (tabell 5.3.2). For arbeidsledige er den på 36 prosent og for de under utdanning 22 prosent. En av grunnene til at feilen blir svært stor, er at andelen arbeidsledige og studenter er svært liten, men dette kan ikke forklare alt.

Tabell 5.3.2. Målefeil for aktivitetsstatus i arbeidsmarkedsregisteret. AKU-utvalget

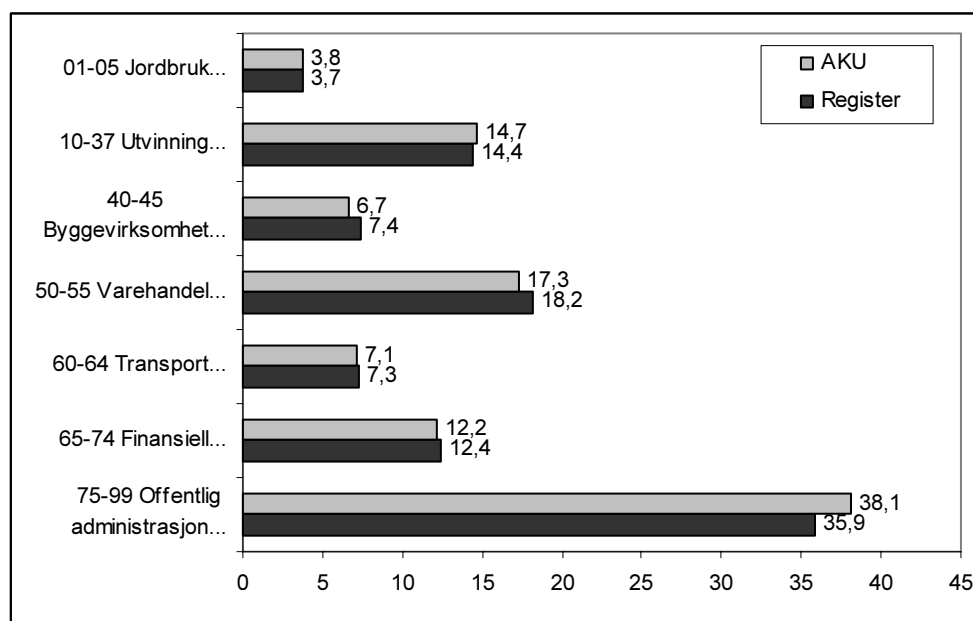
	Antall personer i utvalget	Andel \hat{p}	Systematisk målefeil $\hat{\beta}$		Tilfeldig målefeil $\hat{\tau} / \sqrt{n}$	
			Systematisk målefeil	Relativ målefeil, prosent	Tilfeldig målefeil	Relativ målefeil, prosent
Sysselsatt	15 379	0,714	0,006	0,8	0,002	0,3
Arbeidsledig	338	0,018	-0,006	-36,0	0,001	6,1
Underutdanning	1 126	0,054	-0,012	-22,0	0,001	2,7
Pensjonister	2 972	0,146	0,060	40,8	0,002	1,3
Andre	1 301	0,068	-0,047	-69,5	0,002	3,5

5.4. Sysselsatte etter næring

Hovedkildene for næring er BoF og arbeidstakerregisteret for arbeidsmarkedsregisteret, mens i AKU blir BoF bare brukt som en støtte i klassifikasjonen av næring. De i arbeidsmarkedsregister som ikke har opplysninger om næring fra BoF blir i registersyssetning kodet ut fra regler om utdanning og næring i den øvrige familien. I AKU fins det både hovednæring og binæring, det gjør det ikke i arbeidsmarkedsregisteret. Det kan derfor være i enkelte tilfeller at det blir forskjell mellom AKU og arbeidsmarkedsregisteret, fordi det i registeret er valgt en annen hovednæring enn i AKU.

Andelen sysselsatte etter næring er ganske lik i Arbeidsmarkedsregisteret og AKU, se tabell 5.4.1. Det er noen næringer i registeret som ligger utenfor konfidensintervallet til AKU. Dette er: 40-45 Bygge- og anleggs-virksomhet, kraft- og vannforsyning, 50-55 Varehandel, hotell- og restaurantvirksomhet og 75-99 Offentlig administrasjon og forsvar, helse- og sosialtjenester, undervisning m.m. En av grunnen til denne forskjellen er antakeligvis frafall i AKU, se avsnitt 3.3. Frafallet er lite i næring 75-99 offentlig administrasjon osv. og stort i næring 50-55 Varehandel, hotell- og restaurantvirksomhet og 60-54 Transport og kommunikasjon.

Figur 5.4.1. Sysselsatte etter næring. Arbeidsmarkedsregisteret og AKU. Prosent



Over en tredjedel av alle sysselsatte er sysselsatt i det offentlige (næringene 75-99). Hele 91 prosent av de som er sysselsatt i denne næringen er ifølge registeret også sysselsatt innen disse næringene i AKU. De resterende er hovedsakelig ikke sysselsatt ifølge AKU. (tabell 5.4.2)

Det er omtrent 4 prosent som er sysselsatt innen jordbruk, skogbruk og fiske (næring 01-05). Denne næringen er det dårligst samsvar mellom AKU og register. Av de som er sysselsatt i denne næringen ifølge registeret, er det omtrent 20 prosent av de som ifølge AKU ikke sysselsatt og 5 prosent jobber i det offentlige (næring 75-99).

For de resterende næringen ligger samsvaret mellom AKU og registeret på mellom 80 og 90 prosent. Forskjellen skyldes i hovedsak ulike metoder for å klassifisere personer som sysselsatt og valg av hovedjobb. Det er mellom 5 og 9 prosent av de som er sysselsatt i disse næringene ifølge registeret,

som ikke er sysselsatt ifølge AKU. De resterende er sysselsatt i forskjellige næringer i AKU og registeret. Denne forskjellen kan skyldes valg av hovedjobb enten i AKU eller registeret.

Tabell 5.4.1. Andelen sysselsatte etter næring. Arbeidsmarkedsregisteret og AKU

	Arbeidsmar	AKU 4.	AKU 4. kvartal 2001			
	kedsregister	med	Estimert	Estimert	Standard-	Nedre
	29. okt-	register-	andel	andel	avvik	grense
	4.nov	variabler				Øvre grense
	Andel					
01-05 Jordbruk, skogbruk og fiske	3,7		3,9	3,8	0,15	4,1
10-37 Utvinning av råolje og naturgass, industri og bergverksdrift	14,4		14,9	14,7	0,29	15,2
40-45 Bygge- og anleggsvirksomhet, kraft- og vannforsyning	7,4		6,7	6,7	0,20	7,1
50-55 Varehandel, hotell- og restaurantvirksomhet	18,2		17,6	17,3	0,30	17,9
60-64 Transport og kommunikasjon	7,3		7,0	7,1	0,21	7,6
65-74 Finansiell og forretningsmessig tjenesteyting, eiendomsdrift	12,4		12,6	12,2	0,26	12,7
75-99 Offentlig administrasjon og forsvar, helse- og sosialtjenester, undervisning m.m.	35,9		36,8	38,1	0,39	38,9
Uoppgitt	0,6		0,5	0,2	0,03	0,2

Tabell 5.4.2. Næring i arbeidsmarkedsregisteret og AKU

Arbeidsmarkeds-	AKU									
	I alt	Ikke syssel-	01-05	10-37	40-45	50-55	60-64	65-74	75-99	Uopp-
register		satt	Jordbruk	Utvinning	Bygge-	Varehandel	Transport.	Finansielle	Offentlig	gitt
					virksom-					
					het.					
I alt	100,0	29,2	2,7	10,4	4,7	12,2	5,1	8,6	27,0	0,1
Ikke sysselsatt	100,0	84,2	1,4	1,2	1,1	3,0	0,9	2,0	6,2	0,1
01-05 Jordbruk...	100,0	19,6	69,1	1,9	0,7	1,3	1,3	1,3	4,7	0,0
10-37 Utvinning...	100,0	5,2	0,6	87,8	1,0	1,7	0,5	1,8	0,9	0,4
40-45										
Byggevirksomhet...	100,0	5,6	1,1	2,1	83,4	2,3	1,2	0,6	3,6	0,0
50-55 Varehandel...	100,0	8,7	0,5	1,5	0,9	82,5	0,9	1,8	3,1	0,0
60-64 Transport...	100,0	6,5	0,9	0,9	0,3	1,5	85,9	1,5	2,2	0,2
65-74 Finansielle....	100,0	7,7	0,4	2,2	0,6	3,5	1,3	80,7	3,4	0,3
75-99 Offentlig...	100,0	5,9	0,5	0,4	0,4	0,9	0,3	0,6	91,1	0,0
Uoppgitt	100,0	25,5	0,0	7,1	5,7	10,1	14,1	14,5	23,0	0,0

Det er ingen store skjevheter (systematiske målefeil) i arbeidsmarkedsregisteret for sysselsatte etter næring (tabell 5.4.3). Den relative skjevheten er for alle næringene på under 5 prosent i forhold til AKU. Det er finansielle tjenester (næring 65-74) som har den største relative skjevheten og det er transport (næring 60-64) som har den minste relative skjevheten.

Den tilfeldige målefeilen er også svært liten i registeret og er for alle næringer under 4 prosent i forhold til AKU. Det er jordbruk, skogbruk og fiske (næring 01-05) som har den største relative tilfeldige målefeilen og det er offentlig tjenester (næring 75-99) som har den minste relative tilfeldige målefeilen.

Tabell 5.4.3. Målefeil for sysselsatte etter næring for arbeidsmarkedsregisteret. AKU-utvalget

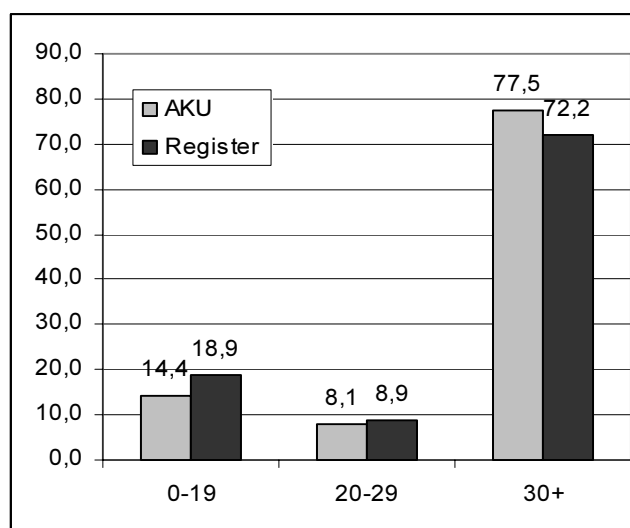
	Antall personer i utvalget	Andel		Systematisk målefeil		Tilfeldig målefeil	
		\hat{p}	$\hat{\beta}$	Relativ målefeil, prosent	$\hat{\tau} / \sqrt{n}$	Relativ målefeil, prosent	
Ikke sysselsatt	6029	0,286	-0,006	-2,1	0,002	0,7	
01-05 Jordbruk...	581	0,028	0,001	2,7	0,001	3,2	
10-37 Utvinning...	2249	0,107	0,003	2,4	0,001	1,0	
40-45							
Byggevirksomhet...	1010	0,048	0,001	1,2	0,001	1,8	
50-55 Varehandel...	2655	0,126	0,003	2,7	0,001	1,1	
60-64 Transport...	1060	0,050	-0,000	-0,7	0,001	1,6	
65-74 Finansielle....	1903	0,090	0,004	4,4	0,001	1,3	
75-99 Offentlig...	5551	0,263	-0,007	-2,8	0,002	0,6	
Uoppgitt	78	0,004	0,003	68,7	0,000	13,0	

5.5. Avtalt/vanlig (gjennomsnittlig) arbeidstid per uke

For 90 prosent av arbeidstakerne i arbeidsmarkedsregisteret innhentes opplysninger om arbeidstid fra arbeidstakerregisteret. For de resterende blir arbeidstiden beregnet ut fra varighet av lønnsforhold og størrelsen på lønn. For selvstendige næringsdrivende er avtalt arbeidstid per uke beregnet ut fra gjennomsnittlig arbeidstid i AKU og næringsinntekt.

Andelen arbeidstakere som jobber 30 timer eller mer i uka er større i AKU enn i registeret. Ifølge AKU jobber 77,5 prosent av arbeidstakerne 30 timer eller mer, mens det ifølge register bare er 72,2 prosent som jobber 30 timer eller mer. Samtidig er det færre i AKU enn i registeret som jobber mellom 0-19 timer i uka. Andelen i registeret ligger utenfor 95 prosent konfidensintervallet for AKU.

Figur 5.5.1. Sysselsatte etter avtalt arbeidstid. Arbeidsmarkedsregister og AKU. Prosent



Tabell 5.5.1. Sysselsatte etter avtalt arbeidstid. Arbeidsmarkedsregister og AKU

Avtalt/vanlig arbeidstid	Arbeidsmarkedsregister 29. okt-4. nov	AKU 4. kvartal 2001 med register- variabler	AKU 4. kvartal 2001			
	Andel	Estimert andel	Estimert andel	Standardavvik	Nedre grense 95 % konfidens- intervall	Øvre grense 95 % konfidens- intervall
0-19 timer	18,9	19,2	14,4	0,3	13,8	14,9
20-29 timer	8,9	8,9	8,1	0,2	7,6	8,5
30 og flere timer	72,2	72,0	77,5	0,3	76,8	78,2

Av de som jobber mellom 0-19 timer ifølge registeret jobber 20 prosent 30 timer eller mer ifølge AKU og i tillegg er 20 prosent ikke sysselsatt ifølge AKU. Blant de som jobber mellom 20-29 timer i uken ifølge registeret er det 37 prosent som jobber mer 30 timer og mer i uka. Blant de som ifølge registeret jobber 30 timer og mer er det 92 som også gjør det ifølge AKU. Grunnen til disse forholdsvis store forskjellene er at arbeidstid er beregnet for mange i arbeidsmarkedsregisteret og da vil ikke tallene stemme på personnivå. En annen grunn til denne forskjellen er valg av hovedjobb i arbeidsmarkedsregisteret og AKU.

Tabell 5.5.2. Avtalt/vanlig arbeidstid i arbeidsmarkedsregister og AKU. Prosent

Arbeidsmarkeds- register	I alt	AKU			Ikke sysselsatt
		0-19	20-29	30+	
I alt	100,0	10,2	5,7	54,9	29,2
0-19 timer	100,0	48,1	11,2	20,0	20,7
20-29 timer	100,0	11,8	44,9	37,2	6,1
30 og flere timer	100,0	2,4	1,8	92,1	3,8
Ikke sysselsatt (og uoppgitt)	100,0	5,6	1,5	8,6	84,2

Skjevheten (systematiske målefeilen) i registeret er stor i forhold AKU for de som er ansatt mindre enn 30 timer. For de som jobber mellom 0-19 timer er den relative skjevheten (systematiske målefeilen) på 26 prosent, mens for de som jobber mellom 20-29 timer er den 9 prosent. For de som jobber 30 timer og mer er den relative systematiske målefeilen på bare 7 prosent. Den tilfeldige målefeilen er forholdsvis liten for alle gruppene.

Tabell 5.5.3. Målefeil avtalt/vanlig arbeidstid for arbeidsmarkedsregisteret. AKU-utvalget

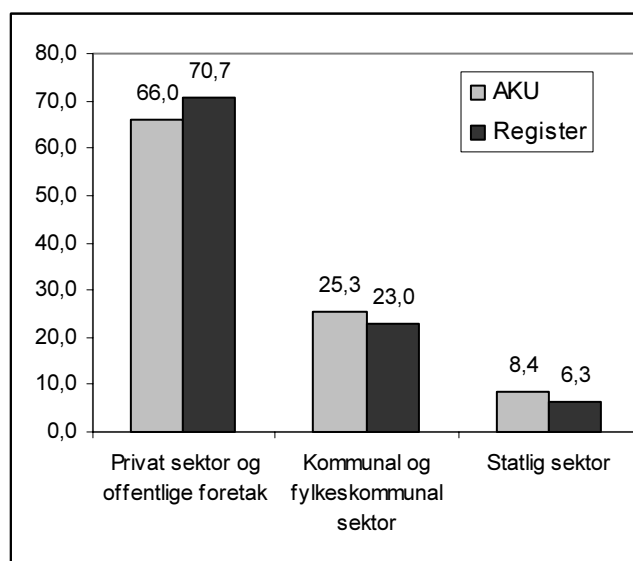
Avtalt arbeidstid	Andel	Systematisk målefeil		Tilfeldig målefeil	
	\hat{p}	$\hat{\beta}$	Relativ målefeil, prosent	$\hat{\tau} / \sqrt{n}$	Relativ målefeil, prosent
0-19 timer	0,137	0,035	25,7	0,002	1,6
20-29 timer	0,063	0,006	9,4	0,002	2,7
30 og flere timer	0,514	-0,035	-6,7	0,002	0,5
Ikke sysselsatt (og uoppgitt)	0,286	-0,006	-2,3	0,002	0,7

5.6. Sektor

66 prosent av arbeidstakerne er sysselsatt i det private ifølge AKU, mens det er 71 prosent ifølge arbeidsmarkedsregisteret. Samtidig er det flere arbeidstakere sysselsatt i statlig, kommunal og fylkeskommunal sektor i AKU enn i arbeidsmarkedsregisteret. Alle tallene fra arbeidsmarkedsregisteret ligger utenfor konfidensintervallet til AKU.

Forskjellen mellom AKU og registeret skyldes ikke oppblåsing i AKU. Hvis vi ser på AKU med registervariabler, er den svært lik fordelingen i registeret. Forskjellen mellom AKU og register må derfor skyldes forskjellen i klassifiseringen av sektor. Grunnen til forskjellen kan være at det er for offentlig foretak kan være feil kategorisert enten i AKU eller arbeidsmarkedsregisteret.

Figur 5.6.1. Sysselsatte etter sektor. Arbeidsmarkedsregisteret og AKU. Prosent



Tabell 5.6.1. Sysselsatte etter sektor

Sektor	Arbeidsmarkedsregisteret 29. okt-4. nov 2001 med registervariabler		AKU 4. kvartal 2001			
	Andel	Estimert andel	Estimert andel	Standardavvik	Nedre grense 95 % konfidensintervall	Øvre grense 95 % konfidensintervall
Statlig forvaltning	6,3	6,1	8,4	0,2	8,0	8,8
Kommunal og fylkeskommunal forvaltning	23,0	23,6	25,3	0,3	24,7	25,9
Privat sektor og offentlig foretak	70,7	70,3	66,3	0,3	65,6	67,0

Det er et forholdsvis godt samsvar for sektor mellom arbeidsmarkedsregisteret og AKU (tabell 5.6.2). For alle sektorene er det mellom 85 og 90 prosent av AKU utvalget som er klassifisert i samme sektor. Det er størst bevegelse mellom privat og ikke sysselsatt.

Tabell 5.6.2. Sysselsatte etter sektor i register og AKU

Arbeidsmarkedsregister	AKU					
	I alt	Ikke sysselsatt	Statlig forvaltning	Kommunal og fylkeskommunal forvaltning I	Privat sektor og offentlig foretak	Uoppgitt
I alt	100,0	29,2	5,9	17,9	46,8	0,2
Ikke sysselsatt	100,0	84,2	1,0	3,3	11,4	0,1
Statlig forvaltning	100,0	4,6	89,7	3,1	2,6	0,0
Kommunal og fylkeskommunal forvaltning	100,0	5,1	1,2	90,6	3,0	0,0
Privat sektor og offentlig foretak	100,0	8,1	3,1	3,1	85,4	0,3

Skjevheten er stor for statlig sektoren, men ikke for privat og kommunal og fylkeskommunal sektor (tabell 5.6.3). For statlig sektor er den relative systematiske målefeilen oppi i 37 prosent, mens den bare er på 6-7 prosent for de andre sektorene.

Grunnen til forskjellen mellom AKU og arbeidsmarkedsregisteret kan være forskjellen i valg av hoved- og biarbeid i disse to kildene. Det skyldes antakeligvis ikke utvalgsfeil eller frafall siden AKU med registervariable er svært lik arbeidsmarkedsregisteret.

Tabell 5.6.3. Målefeil for sektor i arbeidsmarkedsregisteret. AKU-utvalget

Sektor	Andel	Systematisk målefeil		Tilfeldig målefeil	
	\hat{p}	$\hat{\beta}$	Relativ målefeil, prosent	\hat{t} / \sqrt{n}	Relativ målefeil, prosent
Ikke sysselsatt	0,286	-0,006	-2,1	0,002	0,7
Statlig forvaltning	0,043	-0,016	-37,1	0,001	2,5
Kommunal og fylkeskommunal forvaltning	0,169	-0,011	-6,3	0,001	0,8
Privat sektor og offentlig foretak	0,502	0,035	6,9	0,002	0,5

5.7. Antall sysselsatte i bedriften

I arbeidsmarkedsregisteret er opplysninger om antall sysselsatte i en bedriften er laget i hovedsak ut fra opptelling av hvor mange personer som er registrert som sysselsatt i en bedrift ifølge AA-registeret. Det er i tillegg lagt til de som er sysselsatt ifølge LTO og har et koplet bedriftsforhold. Det kan i enkelte tilfeller være vanskelig å fordele de ansatte i et foretak på de forskjellige bedriftene i foretaket.

I AKU er antall sysselsatte i bedriften hentet inn fra intervjuet. Det vil si at intervju objektet er spurt om hvor mange som jobber i samme bedrift. Dette kan for noen være vanskelig da antakeligvis ikke alle arbeidstagere har full oversikt over hvor mange som jobber i bedriften. I AKU er det ikke spurt om eksakt antall sysselsatte, men i hvilken intervall bedriften er i. Det er 8 prosent som ikke har svart på dette spørsmålet i AKU, mens det er 4 prosent uoppgitt i arbeidsmarkedsregisteret. På grunn av den store andelen med uoppgitt i AKU er det vanskelig med noen direkte sammenligning av prosentandelene.

Tabell 5.7.1. Sysselsatte etter antall sysselsatte i bedriften. Prosent

	Arbeidsmarkeds- register 29. okt- 4.nov	AKU 4. kvartal 2001 med register-variabler	AKU 4. kvartal 2001
Antall ansatte	Andel	Estimert andel	Estimert andel
1-10 ansatte	23,0	22,8	22,0
11-19 ansatte	11,2	10,9	12,5
20-49 ansatte	18,3	18,0	17,1
50-99 ansatte	13,1	13,6	10,5
100-199 ansatte	10,9	11,4	8,1
200 og flere ansatte	19,3	19,6	22,0
Uoppgitt	4,3	3,9	7,9

Det er et svært dårlig samsvar i opplysninger om antall sysselsatt i bedriften mellom AKU og arbeidsmarkedsregister (tabell 5.7.2). Samsvaret innen hver kategori ligger på mellom 30 og 70 prosent. Det er størst samsvar for bedrifter på 200 og flere ansatte, mens det dårligste samsvaret er for bedrifter mellom 100-199 ansatte. Det er mange i AKU som har oppgitt en kategori lavere enn det som er registrert i Bedrifts- og foretaksregisteret. Det dårlige samsvaret skyldes antakelig at det er vanskelig for intervjuobjektene i AKU å svare korrekt på dette spørsmålet.

Tabell 5.7.2. Antall sysselsatte i bedriften i register og AKU

	AKU								
Arbeidsmarkeds- register	I alt	ikke sysselsatt	1-10 ansatte	11-19 ansatte	20-49 ansatte	50-99 ansatte	100-199 ansatte	200 og flere ansatte	Uoppgitt
I alt	100,0	29,2	15,6	8,8	12,1	7,4	5,7	15,6	5,6
ikke sysselsatt	100,0	84,2	4,4	1,5	1,9	1,3	0,7	1,9	4,2
1-10 ansatte	100,0	8,2	60,6	7,1	3,7	1,9	1,5	3,0	13,9
11-19 ansatte	100,0	8,1	20,7	48,2	12,0	2,2	1,5	5,3	2,0
20-49 ansatte	100,0	6,4	7,9	15,4	49,5	7,6	2,9	7,5	2,7
50-99 ansatte	100,0	5,1	5,8	5,5	19,0	38,2	9,9	12,7	3,7
100-199 ansatte	100,0	5,9	5,4	4,5	9,1	14,0	33,8	23,2	4,1
200 og flere ansatte	100,0	6,5	3,1	3,0	6,6	4,5	6,4	67,3	2,6
Uoppgitt	100,0	15,8	16,2	6,9	6,9	4,3	6,7	23,6	19,5

For personer sysselsatt i bedrifter på 50-99 og 100-199 ansatte er det en stor skjevhet mellom AKU og arbeidsmarkedsregisteret (tabell 5.7.3). Den relative systematiske målefeilen er på henholdsvis 24 og 30 prosent. Grunnen til denne store skjevheten er at det er mange som har oppgitt i AKU at det er 100-199 sysselsatte i den bedriften de jobber i, mens det i bedriftsregisteret er registrert over 200 ansatte i denne bedriften. Tilsvarende er det mange som ifølge arbeidsmarkedsregisteret er ansatt i en bedrift hvor det er 50-99 ansatte hvor de selv i AKU har oppgitt at det er under 50 ansatte i bedriften. For de andre gruppene av antall ansatte i bedriften ligger den relative målefeilen på mellom 4 og 14 prosent.

Tabell 5.7.3. Målefeil for antall sysselsatt i arbeidsmarkedsregister. AKU-utvalget

	Andel	Systematisk målefeil		Tilfeldig målefeil	
	\hat{p}	$\hat{\beta}$	Relativ målefeil, prosent	$\hat{\tau} / \sqrt{n}$	Relativ målefeil, prosent
ikke sysselsatt	0,286	-0,006	-2,1	0,002	0,7
1-10 ansatte	0,163	0,006	-4,0	0,002	1,5
11-19 ansatte	0,078	-0,011	-13,6	0,002	2,7
20-49 ansatte	0,129	0,007	5,6	0,002	1,9
50-99 ansatte	0,097	0,023	23,6	0,002	2,2
100-199 ansatte	0,081	0,024	29,6	0,002	2,4
200 og flere ansatte	0,140	-0,016	-11,4	0,002	1,6
Uoppgitt	0,028	-0,028	-101,3	0,002	6,7

Den store forskjellen mellom AKU og arbeidsmarkedsregisteret skyldes antakelig at det er svært vanskelig for de som blir intervjuet i AKU å svare nøyaktig på dette. Ansatte i en bedrift har ofte ikke full oversikt over nøyaktig antall ansatte i bedriften. Det vil derfor bli et ganske omtrentlig svar de gir i AKU. Det vil også forekomme feil i arbeidsmarkedsregisteret for antall ansatte i bedriften. Det kan være manglende oppdateringer av AA-registeret eller at fordelingen mellom antall ansatte i foretaket og bedriften ikke er korrekt.

Referanser

Bråten og Fosen. (1998): Definisjon av sysselsetning basert på registerinformasjon. Utarbeidelse av klassifikasjonsrutiner, Notater 98/64, Statistisk sentralbyrå.

Fosen og Solheim. (2000): Avledede variabler i registerstatistikk. To metoder for klassifikasjon av sysselsetningsstatus, Notater 2000/17, Statistisk sentralbyrå.

Bø og Håland. (2002): Dokumentasjon av arbeidskraftsundersøkelsen (AKU), Notater 2002/24, Statistisk sentralbyrå.

De sist utgitte publikasjonene i serien Notater

- 2003/67 H. Tønseth: Kommuneale helseforskjeller -de finnes, men kan de måles? 15s.
- 2003/68 T.M. Normann: Omnibusundersøkelsen mai/juni 2003. Dokumentasjonsrapport. 50s.
- 2003/69 KOSTRA (Kommune- Stat- Rapportering) Rutinebeskrivelse og dokumentasjon. 60s.
- 2003/70 E. Holmøy og B. Strøm: Fordeling av tjenesteproduksjon mellom offentlig og privat sektor i MSG-6. 25s.
- 2003/71 J.K. Dagsvik: Hvordan skal arbeidstilbudseffekter tallfestes? en oversikt over den mikrobaserte arbeidstilbudsforskningen i Statistisk sentralbyrå. 67s.
- 2003/72 A. Steinkellner: Inntektsstatistikk for personer og familier 1999-2001. Dokumentasjon av datagrunnlag og produksjonsprosess. 43s.
- 2003/73 F. Tverå, I. Sagelvmo: Beregning av næringene fiske eget bruk, fiske og fangst og fiskeoppdrett i nasjonalregnskapet. 19s.
- 2003/74 K.H. Grini: Lønnsstatistikk privat sektor 1997-2001. Dokumentasjon av utvalg og beregning av vekter. 36s.
- 2003/75 A.H. Foss: Grafisk revisjon av nøkkeltallene i KOSTRA. 16s.
- 2003/76 K. Hansen: Ideelle organisasjoner i nasjonalregnskapet. 30s.
- 2003/77 E.E. Eibak: Undersøking om foreldrebetaling i barnehagar, august 2003. 46s.
- 2003/78 A.H. Foss: Kvaliteten i husholdningsdelen i Folke- og boligtellingsen 2001. 31s.
- 2003/79 O. Villund: Yrke i Arbeidstakerregisteret. 31s.
- 2003/80 O. Villund: Partielt frafall av yrkesdata i Arbeidstakerregisteret. 18s.
- 2003/81 J.H. Wang: Frafall i konjunkturbarometeret. 45s.
- 2003/82 P. Holmen og K.Lorentzen: Dokumentasjon av etableringen av UT - populasjonen - konsentrasjon om store enheter og stabilitet over tid. 49s.
- 2003/83 T.H. Christensen: Boligprisindeksen. Datagrunnlag og beregningsmetode. 20s.
- 2003/84 G. Dahl: Enslige forsørgere med overgangsstønad. Økonomisk situasjon etter avsluttet stønad. 74s.
- 2003/85 T.M. Normann: Omnibusundersøkelsen august/september 2003. Dokumentasjonsrapport. 36s.
- 2003/86 T. Eika og T. Skjerpen: Hvitevarer 2004. Modell og prognose. 19s.
- 2003/87 S. Blom og B. Lie: Holdningen til innvandrere og innvandring. Spørsmål i SSBs omnibus i august/september 2003. 58s.
- 2003/88 A. Holmøy: Undersøkelse om livsløp, aldring og generasjon (LAG). Dokumentasjonsrapport. 135s.
- 2003/89 Ø. Kleven og E. Wedde: Medieundersøkelsen 2002. Dokumentasjonsrapport. 43s.
- 2003/90 S. Derakhshanfar, S. Lien og C. Nordseth: FD - Trygd. Dokumentasjonsrapport. Barnetrygd. 1996-2002. 44s.
- 2003/91 J. Larsson og K. Telle: Dokumentasjon av DEED . En database over bedriftspesifikke miljødata og økonomiske data for forurensende norske industribedrifter. 16s.
- 2003/92 J.I. Hamre: Undersøkelsen om legemeldt sykefravær. Dokumentasjon av utvalgsplan, trekking og rullering for 2003. 37s.
- 2004/1 A.G. Pedersen: Sammenligning av manuell og auomatisert metode ved koding av dødsårsak. 22s.
- 2004/2 T.M. Köber: Registerbasert sysselsettingsstatistikk for helse og sosialhjelp. 42s.