

Sverre Try

**Datagrunnlag for analyse av arbeidsmarkedsforløp.
Dokumentasjon av registerdata og registerkoblinger til
Kandidatundersøkelsen 2000.**

NIFU skriftserie nr. 6/2003

NIFU – Norsk institutt for studier
av forskning og utdanning
Hegdehaugsveien 31
0352 Oslo

ISSN 0808-4572

Forord

Kandidatundersøkelsen 2000 inngår i første fase av et prosjekt som tar sikte på samle inn longitudinelle data av universitets- og høgskolekandidater som ble uteksaminert våren 2000. Kandidatene skal etter planen følges fra 2000 til 2008. Dette notatet dokumenterer den første fasen i dette arbeidet, herunder resultatene fra en kobling mellom data fra kandidatundersøkelsen og data fra Statistisk sentralbyrås registre.

Prosjektets tittel er *Videreutvikling av kandidatundersøkelsen*, og arbeidet med prosjektet er gjennomført med finansiering fra Norgesnettrådet (nå NOKUT) og Utdannings- og forskningsdepartementet.

Prosjektet er gjennomført og publikasjonen er skrevet av forsker Sverre Try. Ahmed Mohamed i Statistisk sentralbyrå har bistått i forbindelse med koblingen av registerdata. Clara Åse Arnesen, Inger Henaug og Lars Nerdrum har gitt kommentarer til tidligere utkast av rapporten.

Oslo, mars 2003

Petter Aasen
Direktør

Lars Nerdrum
Forskningsleder

Innhold

| | | |
|------------------|---|-----------|
| 1 | Innledning | 7 |
| 2 | Bakgrunn for undersøkelsen..... | 8 |
| 3 | Svarprosent og samtykke..... | 10 |
| 4 | Definisjoner for registerdata og datakilder | 15 |
| 5 | Sammenlikning av en del spørreskjemadata og registerdata | 18 |
| | Referanser..... | 33 |
| Vedlegg 1 | Fordelingstabeller for registerdata | 35 |
| Vedlegg 2 | Spørreskjema Kandidatundersøkelsen 2000 – versjon 1..... | 39 |
| Vedlegg 3 | Spørreskjema Kandidatundersøkelsen 2000 – versjon 2..... | 49 |

1 Innledning

Hovedformålet med prosjektet *Videreutvikling av kandidatundersøkelsen* er å utarbeide et datagrunnlag for analyser av arbeidsmarkedskarrierer blant akademikere etter endt utdanning, ved å følge kandidater som ble uteksaminert fra norske universiteter og høyskoler våren 2000 fra de ble uteksaminert og fram til 2008. Prosjektet vil også være en evaluering av ulike metoder for innsamling av longitudinelle data, ettersom det er lagt opp til to ”konkurrerende” datainnsamlingsprosedyrer. En del av utvalget som deltok i Kandidatundersøkelsen 2000 skal følges opp gjennom nye spørreskjema i 2004 og 2008, og hvor opplysningene fra de nye undersøkelsene kobles til opplysningene fra Kandidatundersøkelsen 2000. Den andre delen av utvalget skal følges gjennom kobling av Statistisk sentralbyrås registerdata til opplysningene som er gitt i Kandidatundersøkelsen 2000. Gjennom kobling av data fra Kandidatundersøkelsen 2000 med registerdata vil vi dessuten kunne vurdere av kvaliteten i opplysningene fra spørreskjema.

Hensikten med dette notatet er å dokumentere første fase av arbeidet med å etablere det longitudinelle datasettet og å presentere registerdataene. I notatet beskrives bakgrunnen for prosjektet nærmere (kapittel 2), og svarprosjenter og andelen kandidater som har samtykket i å delta i de neste fasene av undersøkelsen presenteres (kapittel 3). Videre gis en oversikt over de datasettene som foreløpig er hentet fra SSBs registre (kapittel 4) og til slutt presenteres en sammenlikning av opplysninger som er med både i Kandidatundersøkelsen 2000 og i koblede registerdata fra SSB (kapittel 5). I vedlegg 1 presenteres fordelingen på de koblede registerdataene fra SSB. Hovedresultatene fra Kandidatundersøkelsen 2000 – spørreskjemadata – er tidligere presentert i NIFU (2001), Arnesen (2002) og på NIFUs hjemmesider (www.nifu.no).

2 Bakgrunn for undersøkelsen

En av de sentrale arbeidsoppgavene for NIFUs studier av høyere utdanning er å overvåke hva som skjer med de ferdige kandidatene etter eksamen. Instituttet har siden 1972 gjennomført årlige oppfølgingsundersøkelser av kandidater med høyere utdanning. Kandidatundersøkelsen kartlegger sysselsetting, arbeidsledighet, vansker med å finne passende arbeid, hovedaktivitet, lønnsutvikling osv.

Fram til 1984 omfattet undersøkelsen kandidater fra universiteter, vitenskapelige høyskoler og (de tidligere) distriktshøgskolene. Senere er i enkelte år også andre høyskoleutdanninger inkludert, mens høyere grads kandidater alltid er med i undersøkelsene. Allmennlærere, førskolelærere, sosionomer, barnevernspedagoger, ingeniører, bioingeniører og sykepleiere har ved en eller flere anledninger blitt undersøkt enten som supplement til, eller i stedet for, universitets- og distriktshøgskolekandidatene.

Kandidatundersøkelsene kan grovt deles inn i to forskjellige typer undersøkelser: Ordinære ½-års-undersøkelser og ulike spesialundersøkelser. De ordinære ½-års-undersøkelsene gjennomføres omtrent annethvert år, og noen ganger hvert år. Den første ½-års-undersøkelsen ble gjennomført i 1972, og den foreløpig siste er under planlegging for 2003. Her er sammenliknbarhet mellom undersøkelsene viktig, og ½-års-undersøkelsene legges derfor opp etter det samme designet fra år til år (med revisjoner fra tid til annen). Kandidater fra vårkullene i de grupper som er med i undersøkelsen det enkelte år får tilsendt et spørreskjema et halvt år etter eksamen. Stort sett gjelder dette samtlige kandidater i de enkelte gruppene, men for enkelte utdanningsgrupper har det vært trukket utvalg. Svarprosenten ved disse undersøkelsene har bestandig vært høy, mellom 70 og 80 prosent de siste årene. Disse ½-års-undersøkelsene fanger opp overgangen fra høyere utdanning til arbeidsmarkedet på kort sikt.

Spesialundersøkelser gjennomføres de årene det ikke gjennomføres ordinære ½-års-undersøkelser. Som regel gjennomføres spesialundersøkelser av kandidater flere år etter avsluttet utdanning, for å fange opp sammenhengen mellom utdanning og arbeid over flere år. Disse undersøkelsene har hittil vært retrospektive, det vil si at det stilles tilbakegående spørsmål om yrkesaktivitet, mobilitet og lignende. På denne måten får man kartlagt individuelle karriereforløp på en kostnadseffektiv måte. En innvending mot disse undersøkelsene er imidlertid at det er enkelte forhold det kan være vanskelig å spørre om retrospektivt. Respondentene kan ha problemer med å huske detaljer etter noen år, for eksempel om lønn eller inntekt. Subjektive vurderinger kan også utsettes for erindringsforskyvninger. Dette er en viktig årsak til at NIFU ønsket å prøve ut en annet undersøkelsesdesign for langtidsoppfølging av kandidatene enn vi tradisjonelt har benyttet.

Kandidatundersøkelsen 2000 ble i utgangspunktet gjennomført nesten som en ordinær ½-års-undersøkelse, men med tilrettelegging for oppfølginger i 2004 og 2008 for å kunne følge kandidatenes yrkeskarriere, lønnsforhold, mobilitet, eventuell videreutdanning og så videre.

Undersøkelsen omfatter nesten alle høyere grads kandidater fra norske læresteder, siviløkonomer samt nyutdannede sykepleiere og ingeniører. De to høyskolegruppene ble inkludert etter ønske fra oppdragsgiver (Norgesnettrådet), som var spesielt interessert i karriereutvikling innenfor disse gruppene.

I forhold til en ordinær ½-års-undersøkelse ble følgende tre justeringer foretatt i Kandidatundersøkelsen 2000:

- i) Referansetidspunktet for undersøkelsen ble forskjøvet 2 uker i tid. Normalt gjennomføres ½-års-undersøkelsene midt i november, men for å få tilnærmet samme referansetidspunkt for spørreskjemadata og for koblede registerdata, ble undersøkelsesuka for kandidatundersøkelsen lagt til siste uke i november (27.11-3.12.2000).
- ii) Utvalget ble fordelt i to like store grupper, hvor den ene gruppen skal følges opp med nye spørreskjema i 2004 og 2008, mens den andre gruppen skal følges gjennom kobling av registerdata for 2000, 2004 og 2008. De to gruppene ble fordelt tilfeldig innenfor utdanningsgruppe, lærested og kjønn for å sikre representativitet i begge grupper.
- iii) De to gruppene ble gitt identiske spørreskjema i slutten av november 2000, med følgende unntak: I følgebrevet til de som skal følges opp med nye spørreskjema i 2004 og 2008 ble det orientert om dette, og på slutten av skjemaet spurte vi om respondentene var villige til å delta i de to neste undersøkelsene. I følgebrevet til den gruppen vi ønsket å koble mot registerdata orienterte vi om denne koblingen, og på slutten av spørreskjemaet innhentet vi aktivt samtykke til koblingen. Spørreskjema med følgebrev er inkludert som vedlegg 2 og 3 til dette notatet.

Av hensyn til tidsserier og sammenliknbarhet, er NIFU forsiktig med å eksperimentere med de ordinære ½-års-undersøkelsene. Forskyvningen av referansetidspunktet fører til at kandidatundersøkelsen for 2000 ikke er direkte sammenliknbar med ordinære ½-års-undersøkelser, som for eksempel ble gjennomført i 1999 og 2001. Men tidsforskyvningen er ikke stor, og det er ikke sikkert det betyr så mye. Vi har derfor i enkelte tilfeller valgt å presentere tidsserier hvor undersøkelsen i 2000 er lagt inn sammen med de ordinære ½-års-undersøkelsene, se f.eks. Arnesen (2002) og NIFU (2002). I neste kapittel skal vi se hva endringene i kandidatundersøkelsen for 2002 har betydd for svarprosenten.

3 Svarprosent og samtykke

Kandidatundersøkelsen i 2000 omfatter universitets- og høgskolekandidater med høyere grad, det vil si hovedfagsnivå eller tilsvarende, samt siviløkonomer fra offentlige læresteder, ingeniører og sykepleiere som ble uteksaminert i løpet av vårsemesteret 2000.¹ Alle kandidater med høyere grad og siviløkonomer fra offentlige læresteder², kvinnelige ingeniører og mannlige sykepleiere, samt et utvalg av mannlige ingeniører og kvinnelige sykepleiere inngår i bruttoutvalget.³ Det ble purret to ganger, og svarprosenten totalt sett er 67 prosent, herav 70 prosent blant høyere grads kandidater og 62 prosent blant lavere grads kandidater, jf. tabell 1.

Tabell 1. Populasjon, bruttoutvalg, svarprosent og samtykke til å delta i oppfølging av Kandidatundersøkelsen 2000, etter utdanningsgruppe

| | Populasjon | Bruttoutvalg | Nettutvalg | Svarprosent | Samtykket i videre deltaking | Prosentandel som er med i oppfølgingen |
|---------------------|------------|--------------|------------|-------------|------------------------------|--|
| Høyere grad: | | | | | | |
| Humanister | 489 | 489 | 325 | 66 | 298 | 61 |
| Samfunnsvitere | 609 | 609 | 465 | 76 | 430 | 71 |
| Psykologer | 99 | 99 | 71 | 72 | 65 | 66 |
| Jurister | 503 | 503 | 321 | 64 | 289 | 57 |
| Realister | 448 | 448 | 334 | 75 | 301 | 67 |
| Helsefagkandidater | 167 | 167 | 126 | 75 | 109 | 65 |
| Sivilingeniører | 430 | 430 | 272 | 63 | 238 | 55 |
| Arkitekter | 40 | 40 | 25 | 63 | 18 | 45 |
| Landbrukskandidater | 195 | 195 | 154 | 79 | 144 | 74 |
| Sum høyere grad | 2980 | 2980 | 2093 | 70 | 1892 | 63 |
| Lavere grad: | | | | | | |
| Siviløkonomer | 429 | 429 | 261 | 61 | 230 | 54 |
| Ingeniører | 1426 | 914 | 573 | 63 | 470 | 51 |
| Sykepleiere | 2648 | 729 | 450 | 62 | 381 | 52 |
| Sum lavere grad | 4503 | 2072 | 1284 | 62 | 1081 | 52 |
| Total | 7483 | 5052 | 3377 | 67 | 2973 | 59 |

¹ Legekandidater er ikke med i kandidatundersøkelsene. For øvrig omfatter undersøkelsene alle høyere grads kandidater.

² Det innebærer at BI-kandidater ikke er inkludert i undersøkelsen.

³ På grunn av høye kandidattall i populasjonen av mannlige ingeniører ble det trukket et stratifisert utvalg – innenfor hver skole – av 75 prosent mannlige ingeniørkandidater fra maskinfag og 33 prosent mannlige ingeniørkandidater fra elektrofag, mens alle kandidater fra de andre ingeniørfagene er med i bruttoutvalget. Tilsvarende ble det trukket et stratifisert utvalg av kvinnelige sykepleiere på 20 prosent.

De fleste respondentene som inngår i nettoutvalget samtykket i å delta i oppfølgingen av kandidatundersøkelsene i 2004 og 2008. Andelen av bruttoutvalget som svarte på spørreskjema og som samtykket i å delta i oppfølgingen er 59 prosent, herav 63 prosent blant høyere grads kandidater og 52 prosent blant lavere grads kandidater (jf. tabell 1). Landbrukskandidater og samfunnsvitere er de gruppene som i størst grad har svart på spørreskjema og sagt seg villige til å være med i oppfølgingen av undersøkelsen, med en prosentandel på henholdsvis 74 og 71 prosent. Lavest andel som er med i oppfølgingen finner vi blant arkitekter, med 45 prosent, men dette er en liten utdanningsgruppe og disse bidrar ikke til å trekke den gjennomsnittlige svarprosenten særlig ned. I tillegg ligger oppslutningen blant lavere grads kandidater gjennomgående betydelig lavere enn blant høyere grads kandidater.

Svarprosenten i Kandidatundersøkelsen 2000 ligger noe lavere enn de svarprosentene som det er vanlig å få på slike ½-års-undersøkelser, jf. tabell 2. I perioden 1995-1999 ble det gjennomført fire tilsvarende undersøkelser, og svarprosenten blant høyere grads kandidater varierer fra 79 prosent (i 1997) til 75 prosent (i 1999). Svarprosenten blant høyere grads kandidater i 2000 er signifikant lavere enn den var i alle de fire foregående undersøkelsene. Også blant siviløkonomer er svarprosenten i 2000 lavere enn i de fire foregående undersøkelsene, og forskjellen er signifikant i tre av fire tilfeller. Svarprosenten blant ingeniører ligger på samme nivå i 2000 som i 1999, men er signifikant lavere i 2000 enn i 1997 og 1995.⁴ Tilsvarende ligger svarprosenten blant sykepleierne signifikant lavere i 2000 sammenliknet med i 1995.⁵ Det er også verd å merke seg at svarprosenten er lavere i 2000 enn i 2001, selv om disse forskjellene ikke er signifikante. Alt i alt kan vi altså konkludere med at svarprosenten i Kandidatundersøkelsen 2000 er lavere enn den har vært i tidligere år, menden er ikke signifikant forskjellig fra svarprosenten i 2001.

Den lave svarprosenten i Kandidatundersøkelsen 2000 sammenliknet med tidligere kandidatundersøkelser kan ha sammenheng med at denne undersøkelsen er første del av en panelundersøkelse, og at enkelte kandidater ikke har ønsket å være med på et slikt opplegg. Selv om det var mulig for respondentene å reservere seg mot videre deltaking utover spørreundersøkelsen i 2000, kan omtalen av oppfølgingsundersøkelser isolert sett ha bidratt til å øke frafallet. Men den relativt lave svarprosenten i 2000 må også ses på bakgrunn av en generell trend med fallende svarprosent i utvalgsundersøkelser som følge av sterk vekst i omfanget av kvantitativ datainnsamling (Moe 2001). Således har også kandidatundersøkelsene generelt opplevd økt frafall gjennom andre halvdel av 1990-tallet. Selv om svarprosenten har gått ned i kandidatundersøkelsene i perioden 1995 til 2000 er den imidlertid fortsatt relativt høy sammenliknet med mange andre undersøkelser, og undersøkelsen vurderes som å være representativ for nyutdannede akademikere.

⁴ Ingeniørene var ikke inkludert i kandidatundersøkelsen 1996.

⁵ Sykepleierne var kun med i kandidatundersøkelsene i 1995 og 2000.

Tabell 2. Utvikling i svarprosent for Kandidatundersøkelsen 1995-2001

| | 1995 | 1996 | 1997 | 1999 | 2000 | 2001 |
|-------------------------|------------------|------------------|---------------------|---------------------|------------------|------|
| Høyere grads kandidater | 78 ^{??} | 78 ^{??} | 79 ^{??} | 75 ^{** ??} | 70 ^{**} | 71 |
| Siviløkonomer | 72 ^{??} | 63 ^{**} | 71 ^{** ??} | 68 [?] | 61 [*] | 65 |
| Ingeniører | 80 ^{??} | - | 70 ^{??} | 63 ^{**} | 63 | - |
| Sykepleiere | 69 ^{??} | - | - | - | 62 | - |

Signifikant forskjellig fra svarprosenten året før på ** 1 prosent eller * 5 prosent nivå.

Signifikant forskjellig fra svarprosenten i Kandidatundersøkelsen 2000 på ?? 1 prosent eller ? 5 prosent nivå.

Som nevnt i forrige avsnitt er utvalget delt i to, hvor den ene delen skal følges opp med nye spørreskjema i 2004 og 2008, mens den andre delen av utvalget skal følges gjennom koblinger med dataregistre i SSB. I oversendelsesbrevet til Kandidatundersøkelsen 2000 ble respondentene informert om dette, og aktivt samtykke til å koble informasjonen fra spørreskjema i 2000 med paneldata fra de ulike kildene ble innhentet. Tabell 3 viser hvordan svarprosent og samtykke til videre deltaking/kobling fordeler seg på de to delene av utvalget.

Det er ikke noen stor forskjell i svarprosenten mellom de to delene av utvalget. I den delen av utvalget som skal følges opp med nye spørreskjema i 2004 og 2008 har 68 prosent svart på spørreskjemaet i 2000, og 60 prosent har både svart og samtykket i å delta i den videre oppfølgingen. I den delen av utvalget som skal følges opp gjennom registerkoblinger har 66 prosent svart på spørreskjema og 59 prosent har både svart og samtykket i koblingen. Oppslutningen om undersøkelsen er generelt høyere blant høyere grads kandidater enn blant lavere grads kandidater, og blant samfunnsvitere og helsefagkandidater er det en høyere andel som samtykker i registerkoblingen enn som ønsker å være med spørreskjemaoppfølgingen. For alle andre utdanningsgrupper er det noe større skepsis til registerkoblingen enn det er til å motta nye spørreskjema i 2004 og 2008, til tross for at nye spørreskjema innebærer en merbelastning for respondentene. Dette kan skyldes en generell mistenksomhet til kobling av dataregistre, og en følelse av mangel på kontroll ved å tillate registerkobling. Men forskjellen i oppslutning mellom de to delene av utvalget er ikke særlig stor, og vi skal kanskje ikke legge alt for mye vekt på en slik forklaring. Alt i alt kan det se ut som om den merbelastningen det innebærer å samtykke til videre deltakelse i spørreskjemaautvalget nesten oppveier den usikkerheten enkelte kan føle ved å samtykke i registerkoblingen.

Tabell 3. Populasjon, bruttoutvalg, svarprosent og samtykke til å delta i oppfølging av Kandidatundersøkelsen 2000, etter delutvalg og utdanningsgruppe

| | Bruttoutvalg | Nettoutvalg | Svarprosent | Samtykket i videre deltaking | Prosentandel som er med i oppfølgingen |
|---|--------------|-------------|-------------|------------------------------|--|
| A. SPØRRESKJEMAUTVALG¹: | | | | | |
| Høyere grad: | | | | | |
| Humanister | 242 | 169 | 70 | 153 | 63 |
| Samfunnsvitere | 307 | 222 | 72 | 210 | 68 |
| Psykologer | 50 | 37 | 74 | 34 | 68 |
| Jurister | 250 | 163 | 65 | 151 | 60 |
| Realister | 224 | 165 | 74 | 152 | 68 |
| Helsefagkandidater | 83 | 60 | 72 | 52 | 63 |
| Sivilingeniører | 215 | 144 | 67 | 126 | 59 |
| Arkitekter | 20 | 15 | 75 | 12 | 60 |
| Landbrukskandidater | 94 | 73 | 78 | 70 | 74 |
| Sum høyere grad | 1485 | 1048 | 71 | 960 | 65 |
| Lavere grad: | | | | | |
| Siviløkonomer | 215 | 137 | 64 | 120 | 56 |
| Ingeniører | 463 | 299 | 65 | 241 | 52 |
| Sykepleiere | 364 | 227 | 62 | 195 | 54 |
| Sum lavere grad | 1042 | 663 | 64 | 556 | 53 |
| Sum spørreskjemaavg | 2527 | 1711 | 68 | 1516 | 60 |
| B. REGISTERUTVALG²: | | | | | |
| Høyere grad: | | | | | |
| Humanister | 247 | 156 | 63 | 145 | 59 |
| Samfunnsvitere | 302 | 243 | 80 | 220 | 73 |
| Psykologer | 49 | 34 | 69 | 31 | 63 |
| Jurister | 247 | 158 | 64 | 138 | 56 |
| Realister | 224 | 169 | 75 | 149 | 67 |
| Helsefagkandidater | 84 | 66 | 79 | 57 | 68 |
| Sivilingeniører | 215 | 128 | 60 | 112 | 52 |
| Arkitekter | 20 | 10 | 50 | 6 | 30 |
| Landbrukskandidater | 101 | 81 | 80 | 74 | 73 |
| Sum høyere grad | 1489 | 1045 | 70 | 932 | 63 |
| Lavere grad: | | | | | |
| Siviløkonomer | 214 | 124 | 58 | 110 | 51 |
| Ingeniører | 451 | 274 | 61 | 229 | 51 |
| Sykepleiere | 365 | 223 | 61 | 186 | 51 |
| Sum lavere grad | 1030 | 621 | 60 | 525 | 51 |
| Sum registerutvalg | 2525 | 1666 | 66 | 1457 | 59 |

Note 1: Spørreskjemaavgvalget er den delen av avgvalget som skal følges opp med nye spørreskjema i 2004 og 2008.

Note 2: Registeravgvalget er den delen av avgvalget som skal følges opp gjennom koblinger av registre i SSB i 2004 og 2008.

Den endelige oppslutningen i undersøkelsen varierer en del mellom utdanningsgruppene. En enkel måte å øke svarprosenten på, kunne være å avgrense undersøkelsen til kun å omfatte de utdanningsgruppene hvor svarprosenten er høy. Men dette vil gå på bekostning av antall observasjoner i avgvalget, dersom man ønsker å påvirke svarprosenten utover det neglisjerbare. Dette kan illustreres med eksempel fra registeravgvalget: Dersom vi ekskluderer de utdanningsgruppene som har en svarprosent på under 50, det vil si arkitektene, påvirker ikke

det svarprosenten i det hele tatt, fordi arkitektene er en så liten gruppe (10 personer i registernettutvalget). Men dersom vi ekskluderer alle utdanningsgrupper som har en svarprosent på under 60, det vil si humanister, jurister, sivilingeniører, arkitekter og alle lavere grads kandidater, øker svarprosenten i registerutvalget fra 59 prosent til 70 prosent, men samtidig reduseres nettutvalget fra 1666 observasjoner til 593 observasjoner i registerutvalget. Dette ville innebære et vesentlig tap av data, både i form av antall utdanningsgrupper og antall observasjoner, men gevinsten ville altså være en ganske betydelig økning i svarprosenten for den delen av utvalget vi da står igjen med.

Selv om andelen som samtykker i videre deltakelse i spørreskjemaoppfølgingen er noe høyere enn den er blant de som har samtykket i registerkoblingen, vil den endelige svarprosenten for paneldataene bli mye høyere i registerutvalget enn i spørreskjema utvalget. I prinsippet vil ikke frafallet øke i registerutvalget utover det frafallet vi nå har i Kandidatundersøkelsen 2000, ettersom disse ikke skal motta nye spørreskjema, men kun følges opp gjennom koblinger av registre. I spørreskjema utvalget må vi imidlertid regne med nye frafall i forbindelse med spørreskjemaundersøkelsene i 2004 og i 2008, og det aggregerte frafallet vil kunne bli ganske betydelig. Dersom svarprosenten for spørreskjema utvalget blir det samme i hver av de to neste undersøkelsene som den var i undersøkelsen i 2000, det vil si 68 prosent, blir den aggregerte svarprosenten etter undersøkelsen i 2004 på 41 prosent og den aggregerte svarprosenten etter undersøkelsen i 2008 blir redusert til 27 prosent. Og motsatt: Dersom vi forutsetter en aggregert svarprosent på minst 50 prosent etter undersøkelsen i 2008, må svarprosenten være minst 91 prosent i hver av de to undersøkelsene. Dette er neppe realistisk, og vi må derfor regne med et betydelig frafall i paneldata for spørreskjema utvalget, mens frafallet i registerutvalget ligger fast, dersom vi ser bort fra eventuelle frafall i registrene, på grunn av manglende registrering, eller faktisk frafall i utvalget på grunn av utflytting fra Norge eller dødsfall. I de to neste kapitlene skal vi gå nærmere gjennom datakilder, innhold og kvalitet i registerdataene.

4 Definisjoner for registerdata og datakilder

For de personene i registerutvalget som har samtykket i at opplysningene fra kandidatundersøkelsen kan kobles mot registre i SSB, har vi foretatt den første koblingen mellom data fra Kandidatundersøkelsen 2000 og registerdata. Uttaket for registeropplysningene gjelder i hovedsak for siste uke i november 2000, eller nærmeste registreringstidspunkt til denne uka, med unntak for inntektsopplysningene som refererer til året 2000. Tilsvarende koblinger skal etter planen gjennomføres for samme tidspunkt/perioder i 2004 og 2008. I dette avsnittet presenteres variabeldefinisjoner og datakilder for registerdataene. Fordelingen på ulike variabler følger i vedlegg 1 til notatet.

Følgende variabler er hentet fra Statistisk sentralbyrås personregistre:

Bostedskommune – Registrert bostedskommune per 30.11.2000. Kommunene følger standard kommunekoder per 1.1.2000. Kilde: Personregister fra befolkningsystemet.

Sivilstatus – Registrert ekteskapelig status/sivilstatus per 30.11.2000. Kilde: Personregister fra befolkningsystemet. Klassifikasjon:

- 1 = ugift
- 2 = gift
- 3 = enke/enkemann
- 4 = skilt
- 5 = separert
- 6 = registrert partner

Barn - Antall biologiske barn under 18 år per 31.12.2000. Kilde: Personregister fra befolkningsystemet.

Høyeste fullførte utdanning –Høyeste fullførte utdanning til og med 30.09.2000 klassifisert på NUS2000 seks-siffernivå (standard kodeliste for Norges utdanningsstatistikk, jf. SSB (2001)). Kilde: System for persondata (kobling/kombinasjon av ulike personregistre i SSB) med utdanningsstatistikken som primærkilde (Befolkningens høyeste utdanning, BHU).

Igangværende utdanning – Igangværende utdanning per referansedato (28. november 2000) etter NUS2000 seks-siffernivå (standard kodeliste for Norges utdanningsstatistikk). Kilde: System for persondata (kobling/kombinasjon av ulike personregistre i SSB) med utdanningsstatistikken som primærkilde (register over igangværende utdanning).

Heltids- eller deltidsstudent - Dikotom variabel for om igangværende utdanning per referansedato (28. november 2000) tas på heltid eller deltid. Kvaliteten på denne variabelen er rimelig bra for høgskolestudenter, men ikke for universitetsstudenter der kvaliteten avhenger av opplysningen som studentene selv oppgir når de melder seg til eksamen. Kilde: System for persondata (kobling/kombinasjon av ulike personregistre i SSB) med utdanningsstatistikken som primærkilde (register over igangværende utdanning).

Klassifikasjon:

- 1 = heltid
- 2 = deltid

Sysselsatt eller ikke sysselsatt - Dikotom variabel for sysselsatte (sysselsatte = arbeidstakere + selvstendig næringsdrivende). Det ligger en rekke administrative registre til grunn for klassifiseringen av en persons sysselsettingsstatus, herunder arbeidsgiver/arbeidstakerregisteret som den viktigste kilden for arbeidstakere, med referansedato 28.11.2000. I tillegg hentes det informasjon fra blant annet lønns- og trekkoppgaveregisteret samt enkelte offentlige lønnsregistre. Personer i verneplikt eller siviltjeneste regnes ikke automatisk som sysselsatte, dersom de ellers ikke er registrert som arbeidstaker eller selvstendig næringsdrivende. Kilde: System for persondata (kobling/kombinasjon av ulike personregistre i SSB). Klassifikasjon:

- 1 = sysselsatt
- 2 = ikke sysselsatt

Selvstendig næringsdrivende - Dikotom variabel for selvstendig næringsdrivende per 28.11.2000. Kilde: System for persondata (kobling/kombinasjon av ulike personregistre i SSB). Klassifikasjon:

- 1 = selvstendig næringsdrivende
- 2 = ellers

Arbeidstid - Arbeidstid refererer til avtalt/vanlig arbeidstid for lønnstakere (i hovedarbeidsforholdet for arbeidstakere med mer enn ett arbeidsforhold). Referansedato for arbeidsgiver/arbeidstakerregisteret, som er primærkilden for arbeidstid til lønnstakere, er 28.11.2000. Kilde: System for persondata (kobling/kombinasjon av ulike personregistre i SSB). Klassifikasjon:

- 1 = 0-19 timer
- 2 = 20-29 timer
- 3 = 30 timer eller mer

Arbeidskommune - Arbeidskommune (med utgangspunkt i hovedarbeidsforholdet) per 28.11.2000. (Kommunene følger standard kommunekoder per 1.1.2000.) Kilde: System for persondata (kobling/kombinasjon av ulike personregistre i SSB).

Sektor - Foretakets institusjonelle sektor er definert med utgangspunkt i (hoved)arbeidsforholdet for arbeidstakere per 28.11.2000. Kilde: System for persondata (kobling/kombinasjon av ulike personregistre i SSB).

Startdato for arbeidsforhold - Startdato for nåværende arbeidsforhold, gjelder kun for arbeidstakere. Kilde: System for persondata (kobling/kombinasjon av ulike personregistre i SSB).

Antall ansatte i bedriften - Antall ansatte i bedriften hvor arbeidstakere har sitt (hoved)arbeidsforhold per referansedato 28.11.2000. Kilde: Enhetsregister, bedrift.

Registrert arbeidsledig - Referansedato 28.11.2000. Dikotom variabel for registrerte arbeidsledige. Kilde: SOFA-søkerregisteret. Klassifikasjon:

- 1 = helt arbeidsledig
- 2 = ellers

Start inneværende ledighetsperiode - Registreringsdato (som arbeidssøker/arbeidsledig) for start på ledighetsperioden. Kilde: SOFA-søkerregisteret.

Deltaker på arbeidsmarkedstiltak - Dikotom variabel for deltakere på arbeidsmarkedstiltak. Referansedato 28.11.2000. Kilde: SOFA-søkerregisteret. Klassifikasjon:

- 1 = deltaker på ordinært arbeidsmarkedstiltak
- 2 = ellers

Type arbeidsmarkedstiltak - Inndeling i deltakelse på ulike arbeidsmarkedstiltak per 28.11.2000. Kilde: SOFA-søkerregisteret. Klassifikasjon:

- 0 = helt arbeidsledig
- 1 = offentlige sysselsettingstiltak
- 2 = lønnstilskudd
- 3 = kvalifiseringstiltak (AMO)
- 4 = vikarplasser
- 5 = andre ordinære tiltak
- 6 = yrkeshemmet i ordinære tiltak
- 7 = yrkeshemmet i tiltak for yrkeshe mmede (erstatte attføring)

Start inneværende tiltaksperiode - Registreringsdato for start på eventuell tiltaksperiode. Kilde: SOFA-søkerregisteret.

Dagpengemottaker - Dikotom variabel for dagpengemottakere per 28.11.2000. Kilde: SOFA-søkerregisteret. Klassifikasjon:

- 1 = mottar dagpenger
- 2 = mottar ikke dagpenger

Vernepliktstjeneste - Dikotom variabel militær- eller siviltjeneste per referansedato per 28.11.2000. Kilde: Register for vernepliktige og siviltjenestearbeidere. Klassifikasjon:

- 1 = i sivil/militærtjeneste
- 2 = ellers

Gjenværende studielån - Studiegjeld per 31.12.2000 i kroner. Kilde: Data fra Statens lånekasse for utdanning.

Pensjonsgivende inntekt - Sum pensjonsgivende inntekt for 2000 i kroner. Referansedato 31.12.2000. Kilde: Likningsregisteret.

5 Sammenlikning av en del spørreskjemadata og registerdata

I dette kapitlet skal vi sammenlikne variabler som inngår både i Kandidatundersøkelsen 2000 og i de koblede registerdataene fra SSB og som det er naturlig å sammenlikne med hverandre. Det innebærer at vi ikke vil gå gjennom alle registerdata i dette kapitlet. Noen registervariabler er ikke sammenliknbare med variablene fra kandidatundersøkelsen. Inntektsvariabelen fra registeret, *pensjonsgivende inntekt*, har for eksempel et annet inntektsbegrep enn inntektsvariabelen i kandidatundersøkelsen, *brutto ordinær månedslønn*, og det gir liten mening å sammenlikne de to variablene. I tillegg er det enkelte registervariabler som ikke har noe motstykke i kandidatundersøkelsen i det hele tatt, som for eksempel *gjenværende studielån*. Fordelingen på registervariablene er presentert i vedlegg 1, mens resultater fra Kandidatundersøkelsen 2000 er presentert i NIFU (2001) og Arnesen (2002).

Det er i utgangspunktet de samme personene som inngår i alle data, det vil si de 1457 respondentene som aksepterte at svarene fra kandidatundersøkelsen ble koblet til registerdata. En slik sammenlikning kan fungere som en vurdering av kvaliteten på de ulike datakildene og påliteligheten i svarene som kandidatene selv oppgir i kandidatundersøkelsen. Generelt er selvrapporterte opplysninger til forskningsformål vurdert til å være relativt presise og de er i liten grad påvirket av insentiver til å påvirke resultatene. Selvrapporterte opplysninger er mer presise jo enklere spørsmålene er og jo lettere tilgjengelig svaret er for respondentene, jf. Sawyer m.fl. (1988) for en rekke referanser. Gjennomgangen i dette kapitlet kan gi en indikasjon på kvaliteten i de selvrapporterte opplysningene i kandidatundersøkelsen. Eventuelle avvik mellom opplysningene i kandidatundersøkelsen og i SSBs registre behøver imidlertid ikke å bety at det er opplysningene i kandidatundersøkelsen som er feil, og for mange variabler kan definisjonsforskjeller innebære at det oppstår naturlige mellom de to datakildene. Vi redegjør for disse forskjellene nedenfor.

Bostedskommune

I kandidatundersøkelsen bes respondentene om å oppgi bostedskommune ved intervju tidspunktet, det vil si i månedsskiftet november/desember 2000, jf. spørsmål 3 i vedlegg 2 eller vedlegg 3. Fra SSB har vi fått oppgitt registrert bostedskommune ved utgangen av november 2000, slik dette er registrert i folkeregisteret. Her kan vi forvente avvikende opplysninger på de to variablene av flere grunner. For det første er det ikke presisert nærmere i kandidatundersøkelsen om dette gjelder faktisk eller registrert bosted. Noen respondenter kan derfor tenkes å oppgi den kommunen de er registrert bosatt i, andre kan oppgi den de faktisk bor i. I tillegg kan det være forskjell mellom oppgitt bostedskommune og registrert bostedskommune av flere grunner:

- Studenter skal som hovedregel ikke melde flytting, og kandidater som fortsetter å studere, kan derfor fortsatt stå registrert med en annen bostedskommune enn den hvor de faktisk bor, om de fortsetter å studere.

- Nyutdannede akademikere er en gruppe hvor vi kan forvente relativt stor geografisk mobilitet, og det kan ta noe tid før flyttemelding blir sendt, selv om flyttingen etter regelen skal meldes til folkeregisteret innen åtte dager.
- I overgangen fra utdanningen til arbeid kan det forekomme en del midlertidig opphold, for eksempel i forbindelse med kortvarige engasjementer eller videreutdanninger, hvor kandidatene faktisk ikke skal melde flytting. Midlertidige opphold av under seks måneders varighet skal ikke normalt ikke registreres.

Til tross for disse årsakene til at bostedskommune kan være forskjellig for de to variablene, oppgir faktisk 73 prosent av kandidatene den samme bostedskommunen i kandidatundersøkelsen som den de står registrert med i folkeregisteret. Dette må kunne tolkes som en akseptabel diskrepans mellom de to datakildene, gitt forskjellene i variabeldefinisjon.

Sivilstatus

Kandidatundersøkelsen kartlegger kandidatenes sivilstatus ved intervju tidspunktet (månedsskiftet november/desember 2000, jf. spørsmål 45 i vedlegg 2 eller vedlegg 3). Her er det tre alternativer respondentene kan krysse av på: i) Ugift; ii) Gift/samboende; iii) Separert/ skilt eller enke/enkemann. Fra SSBs registerdata har vi fått koblet til registrert ekteskapielig status ved utgangen av november 2000, etter en seksdeling som framgår av definisjonen i kapittel 4. De to variablene for sivilstatus opererer dermed med samme referansetidspunkt, men det er en viktig forskjell i definisjonen. Registrene i SSB fanger ikke opp samboere, og disse blir dermed stående som ugifte så lenge de ikke inngår ekteskap eller registrert partnerskap.

Tabell 4 viser den bivariate fordelingen i sivilstatus fra de to datakildene, hvor SSBs registerdata er slått sammen til tre grupper. Det er et nesten perfekt samsvar mellom sammenliknbare kategorier i de to datakildene. Alle som oppgir at de er ugift i kandidatundersøkelsen står også registrert som ugift i SSBs registre. Også alle, så nær som en, av de som oppgir å være førgift (separert/skilt/enke/enkemann) i kandidatundersøkelsen står registrert som førgift i SSBs registre. Når det gjelder den tredje kategorien, som omhandler gifte og (delvis) samboende, er det naturlig nok store forskjeller mellom de to datakildene, fordi samboende regnes med i kandidatundersøkelsen men ikke i SSBs registre. Blant de som oppgir å være gift eller samboende i kandidatundersøkelsen, står nesten 40 prosent oppført som gift/registrert partner også hos SSB (300 av 774). De resterende 60 prosent kan vi da regne med er samboere, noe som ikke fanges opp i SSBs registre.

Tabell 4. Sivilstatus oppgitt i kandidatundersøkelsen og registrert i SSB

| | Kandidatundersøkelsen: Oppgitt sivilstatus | | | Uoppgitt |
|---|--|----------------|----------------------------------|----------|
| | Ugift | Gift/samboende | Separert/skilt/ enke/enkemann | |
| <u>Registerdata: Registrert sivilstatus</u> | | | | |
| Ugift | 645 | 459 | 1 | 5 |
| Gift/registrert partner | | 300 | | 1 |
| Separert/skilt/enke/enkemann | | 15 | 25 | |

Barn

I Kandidatundersøkelsen bes respondentene oppgi hvor mange barn de har omsorgsansvar for, og yngste barns alder (se spørsmål 8 i vedlegg 2 eller vedlegg 3). Kandidatundersøkelsen er sendt ut i begynnelsen av desember 2000, og opplysningene om antall barn refererer til denne perioden. Fra registerdataene har vi informasjon om antall biologiske barn under 18 år per 31.12.2000. Re feransetidspunktet for de to variablene er følgelig ikke nøyaktig det samme, men forskjellen i tid er liten, omlag en måned, og har antakeligvis liten betydning. De to variablene har heller ikke helt identisk definisjon, ettersom det kan være forskjell på å ha biologiske barn og å ha omsorgsansvar for barn. Vi vil allikevel forvente stort samsvar mellom de to variablene. Tabell 5 viser den bivarierte fordelingen, hvor vi har satt verdien til null i opplysningene fra kandidatundersøkelsen når yngste barn er over 18 år.

Blant de 1451 personene som SSB har registrert antall biologiske barn for, oppgir 1405 personer å ha omsorgsansvar for samme antall barn i kandidatundersøkelsen, tilsvarende et sammenfall på 97 prosent mellom de to variablene. Personer med samme verdi på de to variablene framgår av hoveddiagonalen i tabell 5, markert med uthevet skrift. Det må betraktes å være et meget stort samsvar mellom de to variablene. I tillegg er det seks personer hvor vi ikke har registrert antall biologiske barn, men hvor vi kjenner antall barn fra kandidatundersøkelsen.

Tabell 5. Antall barn oppgitt i kandidatundersøkelsen og registrert i SSB

| | Kandidatundersøkelsen: Omsorgsansvar for antall barn | | | | | |
|---|--|------------|-----------|-----------|----------|----------|
| | 0 | 1 | 2 | 3 | 4 | 5 |
| <u>Registerdata: Antall biologiske barn</u> | | | | | | |
| 0 | 1149 | 10 | 2 | | | |
| 1 | 5 | 145 | 15 | 3 | 2 | |
| 2 | | 2 | 88 | 3 | 1 | |
| 3 | | | 2 | 20 | | |
| 4 | | | | | 2 | 1 |
| 5 | | | | | | 1 |
| Uoppgitt | 4 | 2 | | | | |

Høyeste fullførte utdanning

NIFU innhenter opplysninger om uteksaminerte kandidater direkte fra de enkelte lærestedene, og sender ut spørreskjema til kandidatene basert på disse opplysningene. Det er opplysningene fra lærestedene, kombinert med de opplysningene respondentene selv gir på spørreskjema (jf. spørsmål 9a-9e i vedlegg 2 eller vedlegg 3), som brukes til å fastslå den utdanningen kandidatene oppnådde i vårsemesteret 2000, klassifisert på NUS2000-sekssiffernivå (jf. SSB 2001). Fra SSB har vi fått oppgitt kandidatenes høyeste fullførte utdanning ved utgangen av september 2000, slik dette er registrert i utdanningsstatistikken, klassifisert på samme måte. Den utdanningen kandidaten avsluttet i vårsemesteret 2000, og

som vi opererer med i kandidatundersøkelsen, skulle dermed være den samme utdanningen som vedkommende står registrert med som den høyeste utdanningen ved utgangen av september samme år i SSB, forutsatt at ikke kandidaten har tatt en utdanning på et enda høyere nivå tidligere. Tabell 6 viser den bivariate fordelingen for de to variablene, hvor utdanningene i hovedsak er fordelt etter hovedgrupper i NUS2000-standarden (to-siffernivå).

Blant de 1416 personene som har registrert en høyeste fullførte utdanning i SSB, står 1367 oppført med samme hovedutdanningsgruppe i kandidatundersøkelsen, tilsvarende et samsvar på 97 prosent. Samsvaret er nesten helt sammenfallende for de fleste utdanningsgrupper, med unntak av kandidater som er oppført som siviløkonomkandidater i kandidatundersøkelsen. I denne gruppen er det en del avvik mellom de to datakildene, ettersom 34 av 110 siviløkonomer i kandidatundersøkelsen står oppført med en annen lavere grads utdanning, eller uten en fullført utdanning, i registerdataene til SSB. Dette misforholdet kan skyldes en forskyvning i semesteret ved Norges Handelshøyskole i forhold til de ordinære vår- og høstsemestrene, som fører til at enkelte personer regnes som kandidater i vårkullet i kandidatundersøkelsen mens de registreres som kandidater i høstkullet samme år i utdanningsstatistikken. Dette kan da føre til at enkelte siviløkonomkandidater står oppført med sin tidligere utdanning som høyeste fullførte i SSBs registre, mens de er oppført som siviløkonomkandidater i kandidatundersøkelsen.

For 40 personer har SSB ikke registrert noen høyeste utdanning, men for disse personene har vi oppgitt hvilken utdanning de fullførte våren 2000 fra kandidatundersøkelsen. Den største gruppen med uoppgitt høyeste utdanning i SSBs registre er siviløkonomkandidater.

Sammenlikningen i tabell 6 er gjort mellom hovedutdanningsgrupper, basert på NUS2000 to-siffernivå. Men begge variablene er i utgangspunktet definert på seks-siffernivå, tilsvarende nesten 700 koder bare for høyere grad. Dersom vi sammenlikner den registrerte utdanningen i SSBs registre med utdanningsopplysningene i kandidatundersøkelsen på dette detaljerte nivået, finner vi at 89 prosent av kandidatene står oppført med nøyaktig samme utdanning i de to datakildene. Når vi aggregerer opp til hovedutdanningsgrupper, som i tabell 5, øker altså dette samsvaret til 97 prosent. Dette skyldes ulik kodepraksis, noe som kan føre til at enkelte kandidater står oppført med utdanningskoder som ligger nær hverandre i de to datakildene, men utdanningene er ikke kodet helt identiske (f.eks. står noen oppført med *724102 cand.paed. hovedfag pedagogikk* det ene stedet, men *724103 cand.paed. profesjonsstudiet pedagogikk* det andre stedet). Slike forskjeller forsvinner på aggregert nivå, og i analytisk sammenheng vil vi stort sett alltid operere med aggregerte utdanningsgrupper. Samsvaret i utdanningsopplysninger i kandidatundersøkelsen og i SSBs registre må derfor kunne betraktes som meget tilfredsstillende.

Tabell 6. Utdanningsnivå oppgitt i kandidatundersøkelsen og registrert i SSB, etter utdanningshovedgrupper

| | Kandidatundersøkelsen: Fullført utdanning vårsemesteret 2000 ¹ | | | | | | | | | |
|--|---|-----------|------------|----------|------------|-----------|-----------|-----------|------------|------------|
| | 1. | 2. | 3. | 4. | 5. | 6. | 7. | 8. | 9. | 10. |
| Registerdata: Høyeste fullførte utdanning | | | | | | | | | | |
| <i>Høyere grad:</i> | | | | | | | | | | |
| 1. Humanistiske og estetiske fag | 145 | | | | | | | | | |
| 2. Lærerutdanninger og pedagogiske fag | | 40 | | | | | | | | |
| 3. Samfunnsfag og juridiske fag | 2 | | 310 | | | | | | | |
| 4. Økonomiske og juridiske fag | | | 1 | 6 | | | | | | |
| 5. Naturvitenskapelige og tekniske fag | | | | 1 | 257 | | | | | |
| 6. Helse-, sosial- og idrettsfag | | | | | | 78 | | | | |
| 7. Primærnæringsfag | | | | | | | 57 | | | |
| <i>Lavere grad:</i> | | | | | | | | | | |
| 8. Siviløkonomer | | | | | | | | 76 | 1 | |
| 9. Ingeniører | | | | 2 | 10 | | | 4 | 217 | |
| 10. Sykepleiere | | | | | | | | 1 | | 181 |
| 11. Andre lavere grads fag | 5 | 2 | | | 1 | 1 | | 15 | 2 | 1 |
| Uoppgitt | 2 | | 4 | | 7 | 1 | | 14 | 9 | 4 |

Note 1: Utdanningshovedgruppene i kandidatundersøkelsen følger samme inndeling som for registerdata, og er i hovedsak basert på NUS2000 to-siffernivå.

Igangværende utdanning / heltids- eller deltidsstudent

Kandidatundersøkelsen inneholder flere spørsmål om igangværende utdanning, jf. spørsmål 41-44 i vedlegg 2 eller vedlegg 3. Her kartlegges studieaktivitet i undersøkelsesuka, det vil si månedsskiftet november-desember 2000. Fra registerdata i SSB har vi fått opplyst om kandidaten står registrert som student ved et norsk lærested på samme tidspunkt, om vedkommende studerer på heltid eller deltid, og hvilken utdanning dette gjelder. Disse opplysningene er hentet fra utdanningsstatistikken.

Selv om referansetidspunktet i prinsippet er det samme for begge de to datakildene, behøver imidlertid ikke opplysningene være de samme i de to datakildene. Kandidater som har registrert seg som student ved et norsk lærested ved begynnelsen av høstsemesteret, vil stort sett fremdeles være registrert som student ved utgangen av november, selv om de ikke lenger studerer og ikke oppgir at de studerer i kandidatundersøkelsen. Kandidater som studerer ved utenlandske læresteder, vil ikke bli fanget opp i registerstatistikken hvis de ikke står registrert ved et norsk lærested, men vil likevel oppgi at de er studenter i kandidatundersøkelsen. I tillegg kommer eventuelle forskyvninger i når lærestedene registrerer at kandidaten avslutter utdanningen og rapporterer dette inn til SSB, som vi også var inne på i omtalen av variabelen over (høyeste fullførte utdanning) som kan føre til at noen kandidater står oppført med igangværende utdanning samtidig som den er registrert som slutført i kandidatundersøkelsen.

Blant de 1457 respondentene som inngår både i kandidatundersøkelsen og i registerdata, er det 1218 som både står registrert som studenter og som oppgir at de studerer i kandidatundersøkelsen, tilsvarende et samsvar på 84 prosent. Av de 239 personene som utgjør avviket mellom de to datakildene, er det 68 personer som oppgir at de studerer i kandidatundersøkelsen, men som ikke er registrert ved et norsk lærested. Tilsvarende er det 171 personer som står registrert som studenter i SSBs utdanningsstatistikk, men som ikke oppgir at de studerer i kandidatundersøkelsen. I den første gruppen, de som sier de studerer blant dem som ikke er registrert som student, studerer 40 prosent ved et utenlandsk lærested, og de inngår derfor ikke i registerstatistikken. I den siste gruppen, de som står registrert som student men som ikke oppgir at de studerer i kandidatundersøkelsen, kan det være noen som registrerte seg som student fordi de var usikre på hva de skulle gjøre høsten 2000, eller fortsatte å studere som et alternativ til arbeidsledighet, men som siden har ombestemt seg eller siden har kommet i arbeid. Det kan dessuten være noen som registrerer seg som student, men som faktisk ikke studerer, for å utnytte ulike rabattordninger.

Tabell 7 viser den bivariate fordelingen mellom SSBs registerdata og kandidatundersøkelsen med hensyn til om kandidatene står oppført som student eller ikke, og om de studerer på heltid eller deltid. De uthevede observasjonene langs hoveddiagonalen utgjør det perfekte samsvaret mellom de to datakildene. Det er imidlertid verd å merke seg at datakvaliteten på om igangværende utdanning er på heltid eller deltid er av varierende kvalitet i registerstatistikken, jf. nærmere omtale ved definisjonen av denne variabelen i kapittel 4.

Tabell 7. Igangværende utdanning oppgitt i kandidatundersøkelsen og registrert i SSB

| | Kandidatundersøkelsen: Beskjeftiget med studier | | |
|---|---|---------------|---------------|
| | Nei | Ja, på heltid | Ja, på deltid |
| Registerdata: Registrert som student | | | |
| Nei | 1021 | 44 | 24 |
| Ja, på heltid | 64 | 105 | 30 |
| Ja, på deltid | 107 | 40 | 22 |

Sysselsatt

Personer som defineres som sysselsatte i kandidatundersøkelsen, omfatter alle som oppfatter seg som hovedsakelig yrkesaktiv eller utførte minst 1 times inntektsgivende arbeid i undersøkelsesuka (27.11-3.12.2000) eller hadde et inntektsgivende arbeid som de var midlertidig borte fra i undersøkelsesuka (jf. spørsmål 15-17 i vedlegg 2 eller 3). Som sysselsatte regnes også de som var i enkelte arbeidsmarkeds-/sysselsettingstiltak (jf. spørsmål 18-19). Denne definisjonen ligger nært opp til definisjonen som brukes i SSBs arbeidskraftundersøkelser (AKU-definisjonen), med det unntaket at vernepliktige ikke automatisk regnes som sysselsatte i kandidatundersøkelsen dersom de ikke tilfredsstillter kriteriene forøvrig. Sysselsettingsvariabelen fra SSBs registre har framkommet ved å koble en rekke administrative registre for å komme så nær AKU-definisjonen som mulig, men heller ikke her er vernepliktige automatisk regnet med, jf. nærmere omtale i forrige kapittel.

Tabell 8 viser den bivariate sysselsettingsfordelingen mellom de to datakildene. I alt 1235 av 1457 personer kategoriseres på samme måte, dvs. et samsvar på 85 prosent. Nesten 12 prosent i utvalget (171 personer) oppgir i kandidatundersøkelsen at de var sysselsatte, men de er ikke registrert som sysselsatte i SSBs registre. Dette kan skyldes en underrapportering i arbeidsgiver/arbeidstakerregisteret, f.eks. når det gjelder kortvarige midlertidige jobber. I denne gruppen oppgir 59 prosent at de er i midlertidige eller tilfeldige jobber. Registeret behøver heller ikke å være helt oppdatert i forhold til nylig etablerte arbeidstakerforhold. Blant kandidater som regnes som sysselsatte i kandidatundersøkelsen men ikke i registerstatistikken, er det 40 prosent som oppgir at de begynte i denne stillingen for inntil 3 måneder siden.

Det er en relativt liten andel (3 ½ prosent) i utvalget som ikke regnes som sysselsatt i kandidatundersøkelsen men som er registrert som sysselsatt i registeret. Også her kan en del av avviket skyldes manglende oppdatering i arbeidsgiver/arbeidstakerregisteret, ved at nylig avsluttede arbeidstakerforhold ennå ikke er registrert. I denne gruppen oppgir 40 prosent av kandidatene at de hadde arbeid ved siden av studiene, og 30 prosent oppgir at de har vært i arbeid etter avsluttet eksamen, dvs. i løpet av det siste halve året. Men her er tallene små, og resultatene må tolkes med forsiktighet.

Tabell 8. Sysselsatte i kandidatundersøkelsen og registrert i SSB

| | Kandidatundersøkelsen: | |
|----------------------|------------------------|-----------------|
| | Sysselsatt | Ikke sysselsatt |
| <u>Registerdata:</u> | | |
| Sysselsatt | 1068 | 51 |
| Ikke sysselsatt | 171 | 167 |

Selvstendig næringsdrivende

Sysselsatte personer bes oppgi hva slags ansettelsesforhold de har i kandidatundersøkelsen, hvor selvstendig næringsdrivende er en svarkategori (spørsmål 29). Fra SSBs register har vi fått en dikotom variabel for samme sysselsettingsforhold, og begge datakildene refererer til månedsskiftet november/desember 2000. Tabell 9 viser fordelingen mellom de to datakildene, og 1436 personer kategoriseres her på samme måte, tilsvarende et samsvar på 99 prosent. Det store samsvaret skyldes imidlertid at de aller fleste kandidatene ikke regnes som selvstendig næringsdrivende i noen av datakildene, mens det bare er 16 personer som regnes som selvstendig næringsdrivende begge steder. I kandidatundersøkelsen er det i tillegg like mange som oppgir at de er selvstendig næringsdrivende, men som ikke framstår slik i registerstatistikken.

Tabell 9. Selvstendig næringsdrivende i kandidatundersøkelsen og registrert i SSB

| | Kandidatundersøkelsen: | |
|----------------------------------|-----------------------------|----------------------------------|
| | Selvstendig næringsdrivende | Ikke selvstendig næringsdrivende |
| <u>Registerdata:</u> | | |
| Selvstendig næringsdrivende | 16 | 5 |
| Ikke selvstendig næringsdrivende | 16 | 1420 |

Arbeidstid

I kandidatundersøkelsen bes kandidater som er i arbeid i undersøkelsesuka, om å oppgi ukentlig arbeidstid. Fra registeret har vi fått oppgitt avtalt/vanlig arbeidstid for lønnstakere (i hovedarbeidsforholdet). Begge datakildene refererer igjen til månedsskiftet november/desember 2000. Allikevel kan vi forvente noen ulike opplysninger i de to datakildene, spesielt fordi registeret opererer med avtalt/vanlig arbeidstid, mens det kan være uklart om kandidatundersøkelsen opererer med avtalt/vanlig arbeidstid eller faktisk arbeidstid i undersøkelsesuka slik spørsmålet er stilt (jf. spørsmål 27 i vedlegg 2 eller vedlegg 3).

Tabell 10 viser fordelingen i de to variablene, hvor kategoriene er inndelt etter stillingsprosent i ukentlig arbeidstid fra kandidatundersøkelsen og avtalt/vanlig antall timer i uka i registeropplysningene. En stillingsprosent på 50 tilsvarer om lag 19 timer i uka (18,75 er 50 prosent av normalarbeidsuka på 37,5), mens 80 prosent tilsvarer nøyaktig 30 timer av en normalarbeidsuke. Samsvaret mellom de to datakildene vil derfor fortsatt følge den uthevede hoveddiagonalen, som i tabell 10 utgjør nesten 90 prosent av observasjonene. Om lag 8

prosent av lønnstakerne oppgir at de jobber lenger i kandidatundersøkelsen enn de står oppført med i registeret. Her kan det være enkelte som har avtalt/normal deltid, men som i den aktuelle undersøkelsesuka faktisk jobbet mer enn normalt. Motsatt er det bare 2 prosent av lønnstakerne som oppgir at de jobbet mindre enn det som framgår av registeret, og dette kan for eksempel skyldes avspasering.

Tabell 10. Ukentlig arbeidstid i kandidatundersøkelsen og avtalt arbeidstid registrert i SSB

| | <u>Kandidatundersøkelsen: Ukentlig arbeidstid</u> | | |
|--|---|---------------|----------------------|
| | 50 prosent eller mindre | 51-79 prosent | 80 prosent eller mer |
| <u>Registerdata: Avtalt arbeidstid</u> | | | |
| 0-19 timer | 36 | 10 | 41 |
| 20-29 timer | 6 | 29 | 30 |
| 30 timer eller mer | 12 | 7 | 831 |

Arbeidskommune

Kandidatundersøkelsen inneholder relativt detaljerte spørsmål om hovdearbeidsforholdet i undersøkelsesuka, og i spørsmål 22 får de sysselsatte spørsmål om i hvilken kommune bedriften/virksomheten ligger. Fra SSB har vi hentet tilsvarende opplysning fra registrene. En samkjøring av opplysningene fra disse to variablene viser at 90 prosent oppgir samme arbeidskommune i kandidatundersøkelsen som det som står i SSBs register. Samsvaret mellom de to statistikk-kildene er altså betydelig bedre når det gjelder arbeidskommune enn det vi fant med hensyn til bostedskommune (se over). Avvikende opplysninger kan blant annet skyldes at arbeidstakere som er sysselsatt i foretak med virksomhet i flere kommuner ikke bestandig vet hvor foretaket har sin hovedtilknytning, og for eksempel oppgir kommunen hvor han/hun selv jobber i kandidatundersøkelsen, mens foretaket står registrert tilhørende en annen kommune i arbeidsgiver-/arbeidstakerregisteret.

Sektor

De sysselsatte i kandidatundersøkelsen blir i spørsmål 23 (jf. vedlegg 2 eller vedlegg 3) bedt om å oppgi hvilken sektor bedriften/virksomheten tilhører. Fra SSBs register har vi fått oppgitt foretakets institusjonelle sektor definert med utgangspunkt i hovedarbeidsforholdet. Begge statistikkildene opererer med månedsskiftet november/desember som referansetidspunkt, og tabell 11 viser den bivariante fordelingen mellom de to variablene blant sysselsatte personer.

Tabell 11. Sektor oppgitt i kandidatundersøkelsen og registrert i SSB

| | Kandidatundersøkelsen: Bedriftens/virksomhetens sektor | | |
|--|--|---------------|----------|
| | Offentlig sektor | Privat sektor | Uoppgitt |
| <u>Registerdata: Foretakets institusjonelle sektor</u> | | | |
| Offentlig sektor ¹ | 465 | 14 | 30 |
| Offentlig forretningsdrift/ off. eide foretak ¹ | 19 | 21 | 3 |
| Privat/personlig eide foretak mv. ² | 27 | 444 | 27 |
| Uoppgitt | 1 | 17 | - |

Note 1: Offentlig forretningsdrift og offentlig eide foretak inkluderer statlig og kommunal forretningsdrift, statlig eide foretak og statsforetak (kode 610-680 i SSBs kode for institusjonell sektor), mens offentlig sektor er stat, fylkeskommune og kommune forøvrig.

Note 2: Private og personlig eide foretak mv. inkluderer finansielle foretak (kode 210-490 i SSBs kode for institusjonell sektor), private og personlig eide foretak, personlig næringsdrivende og ideelle organisasjoner.

Tabell 11 viser et stort samsvar i opplysningene fra de to kildene når det gjelder hovedgruppene av offentlig og privat sektor, markert med uthevet skrift i tabellen. Dersom vi ser bort fra respondenter med uoppgitt sektor i en av de to statistikkildene, er det et samsvar på 96 prosent mellom disse to hovedgruppene. Offentlig forretningsdrift og offentlig eide foretak kan imidlertid komme i en mellomstilling mellom de to hovedsektorene, og her fordeler respondentene seg omtrent likt mellom offentlig og privat sektor i kandidatundersøkelsen. Her kan det være glidende overganger, og spesielt blant dem som jobber i statlig eide foretak, er det mange som kategoriseres i privat sektor i kandidatundersøkelsen. Dette kan for eksempel være statlig heleide eller deleide aksjeselskap. Tilsvarende glidende overganger kan også forekomme blant foretak som står oppført som privat eide selskaper i SSBs register, men hvor det offentlige er deleier, og dette kan være en årsak til at en mindre andel av kandidatene som i følge registeret arbeider i et privat eller personlig eid foretak oppgir i kandidatundersøkelsen at de jobber i offentlig sektor.

En mindre del av de sysselsatte har ikke oppgitt sektor. Andelen uoppgitte er 6 prosent i kandidatundersøkelsen og 2 prosent i registeret, men ingen respondenter står oppført som uoppgitt begge steder, og ved å kombinere opplysningene fra de to variablene skulle det således være mulig å fylle ut sektor for alle respondenter etter en nærmere vurdering.

Startdato for arbeidsforhold

Sysselsatte i kandidatundersøkelsen blir også spurt om når de begynte i den stillingen de har på undersøkelsestidspunktet (se spørsmål 25 i vedlegg 2 eller vedlegg 3). Her oppgis starttidspunktet i månedsnummer og årstall. Fra SSBs register har vi fått oppgitt startdato for det registrerte arbeidsforholdet for arbeidstakere ved utgangen av november 2000. Dersom man samkjører disse to variablene, er det et perfekt samsvar på månedsnivå for 70 prosent av arbeidstakerne. Noe av grunnen til avvik mellom de to variablene kan skyldes erindringsforskyvning hos respondentene i kandidatundersøkelsen, ulik behandling av opphold i arbeidsforhold i de to statistikk-kildene eller mindre målefeil:

- Erindringsforskyvning: Det er en del arbeidsforhold som er påbegynt for flere år siden, og 11 prosent av arbeidsforholdene er påbegynt før år 2000 ifølge registerstatistikken. Det kan være urimelig å forvente at respondentene husker nøyaktig når de påbegynte et arbeidsforhold som de har hatt i lang tid, og at det derfor blir oppgitt unøyaktig starttidspunkt i kandidatundersøkelsen.
- Opphold i arbeidsforhold: Personer som går tilbake til en jobb de har hatt tidligere kan ta utgangspunkt i den opprinnelige datoen for tiltredelse i stillingen når de svarer i kandidatundersøkelsen, mens dette registreres som et nytt arbeidsforhold i registerstatistikken.
- Mindre målefeil: Det er en god del arbeidsforhold hvor avviket i starttidspunkt mellom de to statistikk-kildene bare er en måned, og her kan det være en del tilfeldigheter for nøyaktig hvilken måned starttidspunktet registreres i registerstatistikken eller oppgis i kandidatundersøkelsen.

Tabell 12 viser starttidspunkt kategorisert etter om arbeidstakerne begynte i stillingen/arbeidsforholdet før eller etter år 2000, og eventuelt hvilket kvartal i 2000 stillingen/arbeidsforholdet ble påbegynt. Gruppert på denne måten er det et perfekt samsvar langs den uthevede hoveddiagonalen på 80 prosent. Avvik i tabell 12 kan fortsatt skyldes små og tilfeldige forskjeller, som for eksempel at et arbeidsforhold legges til slutten av et kvartal i kandidatundersøkelsen, men registreres i begynnelsen av neste kvartal i registerstatistikken. Avvik som kan tilskrives mindre etterslep i registerstatistikken, vil føre til at arbeidsforholdet registreres senere i registerdata enn i kandidatundersøkelsen og ligger eventuelt rett under hoveddiagonalen i tabell 12. Dessuten kan det være avvikende forståelse og definisjoner av begrepene *stilling* og *arbeidsforhold*, ved at endring i stillingstittel hos samme arbeidsgiver kan bli regnet som ny stilling i kandidatundersøkelsen, mens det registreres som det samme arbeidsforholdet i registerstatistikken. Slike avvik vil føre til at starttidspunktet i stillingen kommer på et senere tidspunkt i kandidatundersøkelsen enn i registerdata og ligger eventuelt over hoveddiagonalen i tabell 12. Det ligger flere observasjoner over enn under hoveddiagonalen i tabell 12, og dette kan være en indikasjon på at ulike definisjoner av begrepene stilling i kandidatundersøkelsen og arbeidsforhold i registerstatistikken har ført til ulike starttidspunkt for de to variablene.

Tabell 12. Starttidspunkt for stilling i kandidatundersøkelsen og for registrert arbeidsforhold i SSB

| | <u>Kandidatundersøkelsen: Starttidspunkt i stillingen</u> | | | | |
|-------------------------------------|---|-----------------|-----------------|-----------------|-----------------|
| | Før 2000 | 1. kvartal 2000 | 2. kvartal 2000 | 3. kvartal 2000 | 4. kvartal 2000 |
| <u>Registerdata: Starttidspunkt</u> | | | | | |
| <u>for arbeidsforhold</u> | | | | | |
| Før 2000 | 81 | 12 | 27 | 24 | 13 |
| 1. kvartal 2000 | 7 | 38 | 5 | 10 | 2 |
| 2. kvartal 2000 | 6 | 3 | 149 | 16 | 8 |
| 3. kvartal 2000 | 12 | 1 | 18 | 431 | 14 |
| 4. kvartal 2000 | 4 | 0 | 6 | 17 | 102 |

Arbeidsledig

Personer som defineres som arbeidsledige i kandidatundersøkelsen omfatter alle som var uten inntektsgivende arbeid (dvs. var ikke-sysselsatt) og som i tillegg oppfyller minst én av følgende betingelser:

- oppfattet seg som hovedsakelig arbeidsledig og hadde søkt arbeid
- oppfattet seg ikke som hovedsakelig yrkesaktiv eller arbeidsledig, men hadde søkt arbeid de siste fire ukene forut for undersøkelsestidspunktet og kunne ha påtatt seg arbeid i undersøkelsesuka
- var i nærmere definerte arbeidsmarkedstiltak (kvalifiseringstiltak hvor deltakerne ikke mottar lønn).

Definisjonen av arbeidsledighet følger således ikke av et enkelt spørsmål i kandidatundersøkelsen, men bestemmes av svarene på spørsmål 15-19 og 34-35 samlet. Dette gir et arbeidsledighetsbegrep i overensstemmelse med ILO-kriteriene (dvs. at arbeidsledige skal være uten inntektsgivende arbeid, skal aktivt ha søkt etter arbeid og kunne ta arbeid) og det ligger relativt nært opp til definisjonen i Statistisk sentralbyrås arbeidskraftundersøkelse (AKU-ledighet). Arbeidskraftundersøkelsen er imidlertid en utvalgsundersøkelse, og derfor har vi gjennom SSB fått koblet til registrert arbeidsledighet fra Aetats register. Registrert arbeidsledighet omfatter arbeidssøkere uten arbeidsinntekt som har meldt seg til arbeidsformidlingen (Aetat).

Både AKUs og Aetats mål på arbeidsledighet er offisielle mål, og det er derfor foretatt mange sammenlikninger mellom de to målene for å analysere avvik, se f.eks. SSB (2002). Mange av de samme årsakene til avvik mellom AKUs og Aetats mål på arbeidsledighet kan også forklare avvik mellom kandidatundersøkelsens og Aetats mål, herunder:

- Kandidatundersøkelsen kan i større grad fange opp arbeidssøkere som ikke har krav på ledighetstrygd og derfor ikke melder seg til Aetat. Mange nyutdannede akademikere står uten krav på dagpenger.
- Personer på arbeidsmarkedstiltak blir konsekvent ikke klassifisert som registrert arbeidsledige, men kan bli klassifisert som arbeidsledige i kandidatundersøkelsen.
- En person som ikke tilfredsstillt aktivitetskravet (ikke aktivt søkt etter jobb) eller disponibilitetskravet (kan ikke ta arbeid) vil normalt ikke bli definert som arbeidsledig i kandidatundersøkelsen, men kan fortsatt bli registrert som arbeidsledig i Aetat.
- Noen registrerte arbeidsledige oppgir i kandidatundersøkelsen å ha småjobber og blir der klassifisert som sysselsatt.

Disse forholdene trekker i ulik retning. De to første momentene fører til at noen kandidater regnes som arbeidsledige i kandidatundersøkelsen men ikke i registeret. De to siste momentene fører til at noen kandidater er registrert som arbeidsledige hos Aetat, men de regnes ikke som arbeidsledige i kandidatundersøkelsen. Tabell 12 viser fordelingen mellom de to statistikk-kildene for den aktuelle referanseperioden, månedsskiftet november/desember 2000.

Tabell 13. Arbeidsledige i kandidatundersøkelsen og registrert i Aetat

| | Kandidatundersøkelsen: | |
|----------------------|------------------------|-------------------|
| | Arbeidsledig | Ikke arbeidsledig |
| <u>Registerdata:</u> | | |
| Arbeidsledig | 19 | 19 |
| Ikke arbeidsledig | 35 | 1384 |

Totalt sett er det stor meget stor grad av samsvar mellom de to variablene i tabell 13, tilsvarende 96 prosent langs den uthevede hoveddiagonalen. Men dette skyldes at de fleste kandidatene ikke kategoriseres som arbeidsledig i noen av statistikk-kildene. Det er relativt stort avvik i kategoriseringen av arbeidsledige personer. Blant de som står registrert som arbeidsledige i Aetats register, regnes bare halvparten som arbeidsledige også i kandidatundersøkelsen. Og tilsvarende er det bare 19 av 54 kandidater som regnes som arbeidsledige i kandidatundersøkelsen som står registrert som arbeidsledig.

Blant de som er registrert som arbeidsledige men ikke regnes som arbeidsledige i kandidatundersøkelsen, oppgir de fleste at de er sysselsatt, og de fleste av disse igjen er sysselsatt på deltid. Blant de som regnes som arbeidsledige i kandidatundersøkelsen, men som ikke er registrert som det i registeret, oppgir de fleste ikke noen tidligere yrkeserfaring i kandidatundersøkelsen, og noen står også registrert som deltakere på arbeidsmarkedstiltak. Alle disse sammenhengene er i overensstemmelse med forklaringene for mulige avvik mellom kandidatundersøkelsens og Aetats mål for arbeidsledigheten i punktene over, men det er svært få observasjoner i hver av de omtalte gruppene, og vi presenterer derfor ikke eksakte tall innenfor de ulike gruppene. På grunn av få observasjoner presenterer vi heller ikke noen tabell for deltakere på arbeidsmarkedstiltak. Det er bare 11 personer som deltar på ordinære arbeidsmarkedstiltak ifølge registeret, og ni av disse kan vi gjenfinne i kandidatundersøkelsen.

Verneplikt

I kandidatundersøkelsen kartlegges deltakelse i verneplikt gjennom spørsmålet om hovedaktivitet i spørsmål 15 (jf. vedlegg 2 eller vedlegg 3). Det innebærer at det er kandidater som omfatter militær- eller siviltjeneste som sin hovedaktivitet i undersøkelsesuka som fanges opp i kandidatundersøkelsen. Fra registeret for vernepliktige og siviltjenestearbeidere har vi fått koblet til de som står registrert i sivil/militærtjeneste på samme tid. Tabell 14 viser fordelingen mellom disse to variablene. Samsvaret er stort mellom de to variablene, men dette skyldes at de aller fleste ikke er i verneplikt. I kandidatundersøkelsen er det 35 personer som oppgir verneplikt som sin hovedaktivitet, og 23 av disse står oppført i registeret.

Tabell 14. Vernepliktige i kandidatundersøkelsen og registrert i SSB

| | Kandidatundersøkelsen: Verneplikt som hovedaktivitet | |
|--|--|-----------------|
| | Verneplikt | Ikke verneplikt |
| <u>Registerdata: Registrert i verneplikt</u> | | |
| Verneplikt | 23 | 2 |
| Ikke verneplikt | 12 | 1420 |

Oppsummering

Denne gjennomgangen har vist stor grad av samsvar mellom opplysningene i kandidatundersøkelsen og tilsvarende opplysninger i registerdata. Det er et nesten perfekt samsvar mellom mange variabler som opererer med sammenliknbare kategorier. I de tilfellene hvor det er avvik mellom de to statistikk-kildene, kan det som regel forklares med ulike definisjoner eller kategorier. Det kan også være forskjell mellom en registrert status på et gitt tidspunkt, slik denne er fanget opp i SSBs registerstatistikk, og en faktisk aktivitet, slik man har forsøkt å fange denne opp i kandidatundersøkelsen. En person som har registrert seg som student høsten 2000 behøver for eksempel ikke å studere i månedskifte november/desember. Og en person som har registrert seg som arbeidsledig hos Aetat tilfredsstillende ikke nødvendigvis kriteriene for å være arbeidsledig, når vi legger strenge ILO-kriterier til grunn.

Der er knyttet spesiell interesse til samsvaret mellom de to kildene når det gjelder hovedkategoriene i arbeidsstyrken, dvs. antall sysselsatte og antall arbeidsledige, fordi dette er forhold som vil stå sentralt i etterfølgende analyser av kandidatenes arbeidsmarkedsforløp. Når det gjelder antall sysselsatte, er det et samsvar på 85 prosent mellom de to datakildene. Det er noen flere som blir kategorisert som sysselsatte i kandidatundersøkelsen enn i registeret, noe som kan skyldes en underrapportering i arbeidsgiver/arbeidstakerregisteret når det gjelder kortvarige midlertidige jobber. I tillegg kan det være at registeret ikke er helt oppdatert i forhold til nylig etablerte arbeidstakerforhold. Når det gjelder antall arbeidsledige, er det et samsvar på 96 prosent, men dette skyldes at de fleste kandidatene ikke regnes som arbeidsledige i noen av datakildene. Blant de få kandidatene som regnes som arbeidsledige i kandidatundersøkelsen eller i registeret, er det relativt stor variasjon i kategoriseringen mellom de to datakildene. Dette skyldes at det er lagt til grunn ulike definisjoner av arbeidsledighet, og det er viktig å være klar over disse forskjellene ved senere sammenlikninger av arbeidsledighet mellom ulike kilder.

De to statistikk-kildene kan komplementere og utfylle hverandre, for eksempel der hvor det er manglende opplysninger for en variabel. I sammenlikningen av data i tabellene 4-14 har vi presentert frekvensene fra kandidatundersøkelsen og registerdata slik de framkommer fra de to statistikk-kildene isolert. Ved å kombinere opplysningene fra de to kildene vil det være mulig å redusere andelen uoppgitt og supplere informasjonen slik at variablene i materialet forbedres ytterligere.

Gjennomgangen kan også illustrere hvilke analysemuligheter registerdata gir. Kandidatene som er med i Kandidatundersøkelsen 2000 skal følges opp i 2004 og 2008. Halvparten av kandidatene skal etter planen følges opp med nye spørreskjema i 2004 og 2008, mens den andre halvparten skal følges opp gjennom kobling til SSBs registre. Dette kapitlet har vist hvilke type data som da kan være aktuelle. Registerdata gir gode opplysninger om kandidatenes aktivitet og kvantitative tilpasning på arbeidsmarkedet, som for eksempel sysselsettingsstatus (sysselsatt, arbeidsledig eller utenfor arbeidsstyrken), hvor mye de eventuelt jobber (arbeidstid) og inntekt. Men registerdata inneholder færre opplysninger om de kvalitative sidene ved arbeidsmarkedstilpasningen, hvor kandidatundersøkelsen kan supplere adferdsspørsmålene med holdningsspørsmål, vurderinger og synspunkter fra respondentene.

Referanser

- Arnesen, C.Å. (2002). Utviklingen i arbeidsmarkedet for nyutdannede akademikere. I C.Å. Arnesen og N. Sandberg (red.), *Utdanning og arbeidsmarked 2002*, s. 76-111. Oslo: Norsk institutt for studier av forskning og utdanning.
- Moe, A. (2001). Reliabilitet og representativitet i større spørreskjemaundersøkelser. *Norsk Statsvitenskapelig Tidsskrift*, 17:1, s.106-112.
- NIFU (2001). *Kandidatundersøkelsen 2000*. Skriftserie 2001:24. Oslo: Norsk institutt for studier av forskning og utdanning.
- NIFU (2002). *Kandidatundersøkelsen 2001*. Skriftserie 2002:25. Oslo: Norsk institutt for studier av forskning og utdanning.
- Sawyer, R., J. Laing og W. Houston (1989). Accuracy of Self-Reported High School Courses and Grades of College-Bound Students. *College and University*, 64:3, s.288-299.
- SSB (2001). *Norsk standard for utdanningsgruppering. Revidert 2000*. Norges offisielle statistikk C 617. Oslo: Statistisk sentralbyrå.
- SSB (2002). Ulike mål på arbeidsledighet – sprikende ledighetstall? *Økonomiske analyser*, 2002:6, s. 17. Oslo: Statistisk sentralbyrå.

Vedlegg 1 Fordelingstabeller for registerdata

Tabell VI. Frekvensfordeling for kategoriske registerdata fra SSB

| | Antall | Prosent (ekskl. uoppgitt) |
|--|-------------|------------------------------|
| Ekteskapelig status: | | |
| Ugift | 1110 | 76,5 |
| Gift/registrert partner | 301 | 20,7 |
| Skilt/separert | 40 | 2,8 |
| Uoppgitt | 6 | - |
| Sum | 1457 | 100 |
| Antall barn: | | |
| Ingen barn | 1161 | 80,0 |
| 1 barn | 170 | 11,7 |
| 2 barn | 94 | 6,5 |
| 3 barn | 22 | 1,5 |
| 4 eller flere barn | 4 | 0,3 |
| Uoppgitt | 6 | - |
| Sum | 1457 | 100 |
| Høyeste fullførte utdanning: | | |
| <i>Høyere grad:</i> | | |
| 1. Humanistiske og estetiske fag | 145 | 10,0 |
| 2. Lærerutdanninger og pedagogiske fag | 40 | 2,7 |
| 3. Samfunnsfag og juridiske fag | 312 | 21,4 |
| 4. Økonomiske og juridiske fag | 7 | 0,5 |
| 5. Naturvitenskapelige og tekniske fag | 258 | 17,7 |
| 6. Helse-, sosial- og idrettsfag | 78 | 5,4 |
| 7. Primærnæringsfag | 58 | 4,0 |
| <i>Lavere grad:</i> | | |
| 8. Siviløkonomer | 77 | 5,3 |
| 9. Ingeniører | 233 | 16,0 |
| 10. Sykepleiere | 182 | 12,5 |
| 11. Andre lavere grads fag | 26 | 1,8 |
| Uoppgitt | 41 | - |
| Sum | 1457 | 100 |
| Igangværende utdanning: | | |
| Studier på heltid | 199 | 13,7 |
| Studier på deltid | 169 | 11,6 |
| Ikke registrert som student | 1089 | 74,7 |
| Sum | 1457 | 100 |
| Sysselsatt: | | |
| Sysselsatt | 1119 | 76,8 |
| Ikke sysselsatt | 338 | 23,2 |
| Sum | 1457 | 100 |

tabell VI fortsetter

| | Antall | Prosent (ekskl. uoppgitt) |
|---|--------|------------------------------|
| Selvstendig næringsdrivende: | | |
| Selvstendig næringsdrivende | 21 | 1,4 |
| Ikke selvstendig næringsdrivende | 1436 | 98,6 |
| Sum | 1457 | 100 |
| Arbeidstid: | | |
| 0-19 timer | 130 | 11,8 |
| 20-29 timer | 69 | 6,3 |
| 30 timer og over | 901 | 81,9 |
| Uoppgitt (inkludert ikke sysselsatte) | 357 | - |
| Sum | 1457 | 100 |
| Sektor: | | |
| Offentlig sektor | 523 | 47,7 |
| Offentlig forretningsdrift / offentlig eide foretak | 45 | 4,1 |
| Privat / personlig eide foretak | 529 | 48,2 |
| Uoppgitt (inkludert ikke sysselsatte) | 360 | - |
| Sum | 1457 | 100 |
| Registrert arbeidsledig: | | |
| Helt arbeidsledig | 38 | 2,6 |
| Ikke helt arbeidsledig | 1419 | 97,4 |
| Sum | 1457 | 100 |
| Deltaker på arbeidsmarkedstiltak: | | |
| Deltaker på ordinært arbeidsmarkedstiltak | 11 | 0,8 |
| Ikke deltaker på ordinært arbeidsmarkedstiltak | 1446 | 99,2 |
| Sum | 1457 | 100 |
| Dagpengemottaker: | | |
| Mottar dagpenger | 15 | 1,0 |
| Mottar ikke dagpenger | 1442 | 99,0 |
| Sum | 1457 | 100 |
| Vernepliktstjeneste: | | |
| I sivil/militærtjeneste | 25 | 1,7 |
| Ellers | 1432 | 98,3 |
| Sum | 1457 | 100 |

Tabell V2. Verdier for numeriske registerdata fra SSB

| | N: Antall observasjoner | Gjennomsnitt | Median | Standardavvik | Min-maks |
|---|----------------------------|----------------------|----------------------|---------------------|----------------------------|
| Antall ansatte i bedriften (ekskluderer alle med verdi 0) | 1091 (1090) | 642 (642) | 97 (97) | 1338 (1338) | 0-6054 (1-6054) |
| Gjenstående studielån (kr) (ekskluderer alle med verdi 0) | 1457 (1252) | 175.831 (204.621) | 190.164 (208.756) | 100.099 (75.934) | 0-380.209 (693-380.209) |
| Pensjonsgivende inntekt (kr) (ekskluderer alle med verdi 0) | 1457 (1412) | 142.354 (146.891) | 136.900 (139.900) | 89.901 (87.596) | 0-809.700 1.200-809.700 |

Vedlegg 2 Spørreskjema Kandidatundersøkelsen 2000 – versjon 1

Spørreskjema som gikk ut til kandidater som ble bedt om tillatelse til å koble data fra kandidatundersøkelsen til registerdata.

Vedlegg 3 Spørreskjema Kandidatundersøkelsen 2000 – versjon 2

Spørreskjema som gikk ut til kandidater som ble bedt om å delta i nye spørreundersøkelser i 2004 og 2008.

