

Evalvering av utbildningsregistret

2006:4

I serien Bakgrundsfakta presenteras bakgrundsmaterial till den statistik som SCB producerar inom området befolkning och välfärd. Det kan röra sig om produktbeskrivningar, metodredovisningar samt olika sammanställningar av statistik som kan ge en överblick och underlätta användandet av statistiken.

Utgivna publikationer från 2001 i serien Bakgrundsfakta till Befolknings- och välfärdsstatistik

- 2001:1 Hälsindex
- 2001:2 The role of official statistics in a democratic society: statistics as an investment-free for users
- 2001:3 Co-operations project "Population Statistics"
- 2001:4 Uppdatering av immigrationsuppgifter med anpassning till den nya RTB-strukturen
- 2001:5 Flergenerationsregistret – The Multi-Generation Register
- 2001:6 Utredning om komplettering och utvidgning av Flergenerationsregistret
- 2001:7 Konferens om välfärdsstatistik den 27 mars 2001
- 2002:1 Komplettering av Flergenerationsregistret – resultat från provregistrering av personakter över avlidna 1947 – 1967
- 2002:2 Ett nytt system för Register över totalbefolkningen. Bättre kvalitet och nya möjligheter – A new population register system. Better quality and new possibilities.
- 2002:3 Slutrapport: Översyn av SCB:s Intäkts- och kostandsundersökning för flerbostadshus. Bilaga: Minnesanteckningar från programmet boende. Byggnad och bebyggelse om IKU den 22 maj 2001.
- 2002:4 Befolkningens boende – idé till årlig sammanställning över nationell och internationell boendestatistik
- 2002:5 Översyn av SCB:s Bostads- och Hyresundersökning (BHU)
- 2002:6 Kvalitetssäkringsprogram av postenkätundersökningar. BV/ENK:s kvalitetssäkringsprogram
- 2003:1 Longitudinellt barnregister
- 2003:2 Översyn av HBS (HUT, HBU, Utgiftsbarometern) inför 2003 års undersökning
- 2003:3 Manuell registrering eller skanning
- 2003:4 BV-avdelningens Utvecklingsplan 2003
- 2003:5 Flergenerationsregistret 2002, En beskrivning av innehåll och kvalitet
- 2003:5.1 Multi-Generation Register 2002, A description of contents and quality
- 2003:6 Barns hälsa – Förslag till utbyggnad av SCB:s årliga statistik om Barn och deras familjer
- 2003:7 Översyn av Outhyrda lägenheter i flerbostadshus 2002 - 2003-09-12
- 2003:8 Studie av bortfallet i 2000-års undersökning av levnadsförhållanden (ULF)
- 2003:9 Alternativa datainsamlingsmetoder för ULF
- 2003:10 Slutrapport från projektet för översyn av produktionssystemet för framställning av tabeller till publikationer
- 2003:11 Ensamstående och sammanboende i folkbokföringen, taxeringen och i verkligheten

Fortsättning på omslagets tredje sida!

Ovannämnda rapporter, liksom övriga SCB-publikationer, kan beställas från:
Statistiska centralbyrån, SCB, Publikationstjänsten, 701 89 ÖREBRO,
telefon 019-17 68 00 eller fax 019-17 64 44.

Du kan också köpa SCB:s publikationer i **Statistikbutiken**:
Karlavägen 100, Stockholm

Bakgrundsfakta

Evalvering av utbildningsregistret

Befolknings- och välfärdsstatistik 2006:4

**Statistiska centralbyrån
2006**

Background Facts

Population and Welfare Statistics 2006:4

Evaluation of the Swedish register of education

Statistics Sweden
2006

Tidigare publicering se omslagets insida
Previous publication listed at the inside of the cover

Producent SCB, avdelningen för befolkning och välfärd
Producer Statistics Sweden, Population and Welfare Department
701 89 Örebro

Förfrågningar Ann-Charlott Larsson +46 19 17 63 14
Inquiries lotta.larsson@scb.se

© 2006 Statistiska centralbyrån

Enligt lagen (1960:729) om upphovsrätt till litterära och konstnärliga verk är det förbjudet att helt eller delvis mångfaldiga innehållet i denna publikation utan medgivande från Statistiska centralbyrån

Any reproduction of the contents of this publication without prior permission from Statistics Sweden is prohibited by the Act of Copyright in Literary and Artistic Works (1960:729).

Om du citerar ur denna publikation, var god uppge källan på följande sätt:
Källa: SCB, Bakgrundsfakta, *Evalvering av utbildningsregistret*

When quoting material from this publication, please state the source as follows:
Source: Statistics Sweden, Background Facts, *Evaluation of the Swedish register of education*

ISSN 1650-4917
ISBN 91-618-1334-6
ISBN 978-91-618-1334-6
URN:NBN:SE:SCB-2006-BE96ST0604_pdf

Printed in Sweden
SCB-tryck, Örebro 2006:06

Förord

Statistiken över befolkningens utbildning bygger på individregistret Befolkningens utbildning – i dagligt tal kallat Utbildningsregistret (UREG). UREG innehåller högsta genomförda utbildning för i Sverige folkbokförda personer i åldrarna 16 till 74 år per den 1 januari varje år. UREG genomgick till version 2001-01-01 omfattande förändringar. Svensk utbildningsnomenklatur, SUN anpassades, genom SUN 2000, till den internationella nomenklaturen ISCED 97 (International Standard Classification of Education). Ett antal nya uppgiftskällor bearbetades och kodades enligt SUN 2000 för att därefter användas som indata till UREG. Detta har medfört att utbildningsnivån i riket enligt UREG har ökat kraftigt främst på nivåerna 'treårig gymnasieutbildning' och 'eftergymnasial utbildning tre år eller längre'.

För både användare och statistikproducenten är det viktigt att få en bild över kvaliteten i statistiken från UREG. Den senaste evalveringen av statistiken från UREG genomfördes i samband med evalveringen av folk- och bostadsräkningen 1990 (FoB90). I och med de stora förändringarna av UREG till version 2001-01-01 behövdes färskare mått på registrets kvalitet än de som beräknades vid evalveringen 1990. Det togs därför beslut om att genomföra en ny evalvering.

Denna evalvering avsåg uppgifterna om högsta genomförda utbildning för individerna i UREG version 2003-01-01 i form av utbildningsnivå och utbildningsinriktning enligt SUN 2000. I evalveringsstudien har jämförelser gjorts mot motsvarande uppgifter i undersökningen om levnadsförhållanden (ULF) 3:e och 4:e kvartalet 2002.

Syftet med en evalvering av UREG är att mäta tillförlitligheten i statistiken från UREG och identifiera eventuella fel som har uppkommit i samband med insamling och bearbetning av uppgifter till registret. Statistikproducenten är dessutom intresserad av att analysera hur felen uppkommer för att därigenom kunna undvika dem i kommande statistik. Kartläggningen kan leda till förbättringar av statistiken.

Arbetet med evalveringsstudien har genomförts i projektform med Ann-Charlott Larsson som projektledare. I projektet deltog Henrik Gustafsson, Hans Odelholm, Ulla Pettersson och Peter Öberg. Hassan Mirza har bidragit med värdefulla kommentarer och synpunkter på rapporten.

Statistiska Centralbyrån i juni 2006

Anna Wilén

Kajsa Swenson

Innehåll

Förord	3
1 Sammanfattning	7
2 Inledning	9
2.1 Bakgrund.....	9
2.1.1 Svensk utbildningsnomenklatur (SUN 2000).....	10
2.2 Syfte.....	11
3 Deltagare i projektet.....	13
4 Behov	14
4.1 Användare	14
4.2 Användningsområden	14
5 Genomförande.....	15
5.1 Bestämma "sanna" koder	16
5.1.1 Tidigare evalveringar.....	16
5.1.2 Skapa evalveringsundersökningen.....	17
5.1.3 Matchning mellan UREG och ULF	18
5.1.4 Granskning	19
5.2 Metoder för att ta fram skattningar	21
5.2.1 Mått på tillförlitligheten i statistiken.....	21
5.2.2 Urvalsdesign	24
5.2.3 Skattningsförfarande.....	25
6 Resultat.....	26
6.1 Tillförlitligheten i statistiken.....	26
6.1.1 Skillnader mellan olika grupper i populationen.....	28
6.2 Orsaker till skillnader mellan Utbildningsregistret och ULF	36
6.2.1 Utbildningsnivå	36
6.2.2 Utbildningsinriktning.....	39
Bilaga 1.....	41
Skattningsförfarande i evalveringsstudien.....	41

1 Sammanfattning

För samtliga personer i evalveringsstudien genomfördes en matchning av utbildningsnivå respektive utbildningsinriktning för personens högsta utbildning mellan UREG och ULF. I de fall som uppgifterna var lika antogs de vara "sanna". De poster som inte matchade på utbildningsnivå eller utbildningsinriktning gick vidare till granskningen. Syftet med granskningen var i första hand att bestämma en "sann" kod för utbildningsnivå och/eller utbildningsinriktning baserad på tillgänglig information från ULF respektive UREG. I vissa fall var det inte möjligt att bestämma en "sann" kod. Orsaker till varför UREG och ULF inte matchade innan granskning sattes också under granskningen.

Andelen rätt klassificerade i UREG har skattats till 85 procent för utbildningsnivå respektive utbildningsinriktning i evalveringsstudien. Andelen rätt klassificerade är högre för eftergymnasial utbildning än för gymnasial- och förgymnasial utbildning. När det finns skillnad i utbildningsnivå mellan UREG och den "riktiga" klassificeringen i evalveringsstudien har utbildningsnivån oftast höjts ett steg i evalveringsstudien. Ett exempel på detta är att felaktigt klassificerade med förgymnasial utbildning i UREG ofta blivit klassificerade till gymnasial utbildning två år eller kortare i evalveringsstudien.

En stor andel av personerna med förgymnasial utbildning har en allmän utbildning, vilket gör att andelen rätt klassificerade bland personer med allmän utbildning i UREG är relativt låg, jämfört med övriga utbildningsinriktningar. De med felaktig klassificering av utbildningsinriktning i UREG är spridda över ett antal olika utbildningsinriktningar enligt "sann" inriktning i evalveringsstudien.

Personer med okänd utbildningsnivå eller utbildningsinriktning i UREG ingår också i felaktiga uppgifter om högsta utbildning. Majoriteten av personerna med okänd utbildningsnivå i UREG skulle egentligen vara klassificerade inom förgymnasial utbildning enligt evalveringsstudien. Drygt hälften av personerna med en okänd utbildningsinriktning i UREG har blivit klassificerade som allmän utbildning i evalveringsstudien. Därefter har ungefär 30 procent blivit klassificerade som teknisk eller samhällsvetenskaplig utbildning.

Skattningarna av nettofel tyder på att antalet personer med förgymnasial utbildning överskattas i UREG medan antalet personer inom samtliga övriga utbildningsnivåer underskattas. Detta mönster fanns även i den tidigare evalveringen av UREG.

De största problemen, enligt skattningarna av fel, tycks finnas vid klassificeringen av gymnasial utbildning tre år respektive eftergymnasial utbildning två år eller kortare. För utbildningsinriktning tyder skattningarna av nettofel på att antalet personer med allmän utbildning överskattas i UREG medan antalet personer inom nästan samtliga övriga utbildningsinriktningar underskattas. Minst problem med tanke på skattningarna av fel tycks finnas vid klassificeringen av pedagogik- och lärarutbildningar.

Det finns skillnader mellan olika grupper av populationen när det gäller felaktiga uppgifter om utbildningsnivå/utbildningsinriktning i UREG. Bland annat fångas utbildningar inom det reguljära utbildningssystemet upp i högre grad än kortare yrkesutbildningar.

De reguljära utbildningarna hämtas in via statistikregister som täcker bland annat grundskola, gymnasium, kommunal vuxenutbildning och högskola. De största grupperna av källor till UREG är Folk- och Bostadsräkningen (FoB) 1990 och reguljära statistikregister. Evalveringsstudien visar att det är större problem med klassificeringen av utbildningsnivå respektive utbildningsinriktning om uppgiften är hämtad från FoB 1990 än om den är hämtad från reguljära statistikregister.

Det finns även mer fel i uppgifterna om utbildningsnivå respektive utbildningsinriktning bland personer födda utomlands än bland personer födda i Sverige. Detta beror delvis på att utländska utbildningar inte fångas upp i lika hög grad som svenska utbildningar i UREG.

Den största skillnaden gällande felen i UREG mellan förvärvsarbetande och ej förvärvsarbetande är att överskattningen av antalet personer med förgymnasial utbildning verkar vara betydligt större bland de förvärvsarbetande.

I ULF finns det särskilda frågor om deltagande i kommunal vuxenutbildning och övrig utbildning omfattande minst två månader på heltid. Vid jämförelser mellan deltagare och ej deltagare i dessa utbildningsformer gällande skattningarna av fel i UREG, kom det fram att det finns mer problem bland dem som har deltagit än bland dem som inte har deltagit. Ett exempel på detta är att överskattningen av antalet personer med förgymnasial utbildning är betydligt större bland dem som har deltagit i någon övrig utbildning än bland dem som inte har deltagit. Det framkom också att det finns mer problem med uppgifterna om utbildningsnivå bland personer som har deltagit i utbildning på universitet/högskola och inte tagit ut examen än bland dem som har examen från universitet/högskola. Underskattningen av antalet personer med eftergymnasial utbildning är betydligt större bland personer som inte tagit ut examen än bland dem som har examen från universitet/högskola.

Den vanligaste anledningen till att personer har olika utbildningsnivå i UREG och ULF, när "sann" nivå i evalveringsstudien är satt enligt ULF, är att utbildning som är angiven i ULF saknas i UREG. Det handlar ofta om utbildningar utanför det reguljära utbildningssystemet, t.ex. längre lärlingsutbildningar eller andra icke nivåspecifiserade utbildningar som exempelvis "Bageri 1 år" och "Snickarlärling 4 år". Sådana utbildningar rapporteras inte till UREG på administrativ väg. En annan brist i UREG gäller 2-åriga gymnasieutbildningar (före 1990) som har blivit angivna i ULF men inte i FoB 90. Det rör sig troligen om ej avslutade utbildningar eftersom de inte finns i gymnasierregistren. När uppgifter till UREG har hämtats från källan "Högskoleregistret, summerade högskolepoäng" saknas ofta poäng i UREG. Summerade högskolepoäng finns endast från 1993 och framåt.

2 Inledning

2.1 Bakgrund

Statistiken över befolkningens utbildning bygger på individregistret Befolkningens utbildning – i dagligt tal kallat Utbildningsregistret (UREG). UREG innehåller högsta genomförda utbildning för i Sverige folkbokförda personer i åldrarna 16 till 74 år per den 1 januari varje år.

UREG har till version 2001-01-01 genomgått omfattande förändringar enligt följande;

- Svensk utbildningsnomenklatur, SUN, har genom nya SUN 2000 anpassats till den internationella nomenklaturen ISCED 97 (International Standard Classification of Education). Till skillnad från gamla SUN består SUN 2000 av två delar; en nivådel och en inriktningsdel som kan användas tillsammans eller var för sig.
- Ett antal nya uppgiftskällor har bearbetats och kodats enligt SUN 2000 och därefter använts som indata till UREG. De nya källorna finns beskrivna nedan;
 - Summerade högskolepoäng fr.o.m. 1993 från högskoleregistret
 - Summerade komvux-poäng fr.o.m. 1988
 - Utländska utbildningar som värderats av Högskoleverket.
 - Socialstyrelsens register över hälso- och sjukvårdspersonal
 - Sökande till högskolan från Universitets- och högskoleregistret
 - Grundmaterial från försök med insamling av samlade betygsdokument i gymnasieskolan från läsåret 98/99.

De stora förändringarna till version 2001-01-01 av UREG har medfört att utbildningsnivån i riket enligt UREG har ökat kraftigt främst på nivåerna 'treårig gymnasieutbildning' och 'eftergymnasial utbildning tre år eller längre'. Utbildningens inriktning har inte förändrats i lika stor omfattning, dock har många enskilda utbildningar klassificerats under ny inriktning till följd av nomenklaturbytet.

En tidigare evalvering av UREG genomfördes i samband med evalveringen av folk- och bostadsräkningen 1990 (FoB90). Då jämfördes utbildningsnivå respektive utbildningsinriktning mellan arbetskraftsundersökningen (AKU) och FoB90. Personer med matchande koder antogs ha en "sann" kod för högsta utbildning. För resterande personer genomfördes granskning samt en telefonuppföljning för att avgöra vilken kod som kunde antas vara "sann". Detta resulterade i en fil med "sanna" koder för högsta utbildning som sedan användes vid jämförelser mot UREG.

Evalveringen av UREG 1991 ledde till några viktiga förändringar i produktionen av registret. Framförallt ändrades den inbördes prioriteringen av källor och poängproduktionen från högskola/universitet tillkom som källa.

I och med de stora förändringarna av UREG till version 2001-01-01 behövs färskare mått på registrets kvalitet än de som beräknades vid evalveringen 1990. Därför togs beslutet att en förstudie skulle genomföras för att utreda hur evalveringen skulle genomföras och vilken källa som var lämpligast att använda som bas för evalveringsundersökningen. Valet stod mellan undersökningarna av levnadsförhållanden 2002 (ULF) och en pilotundersökning om vuxnas lärande som genomfördes under våren 2002.

Efter en analys av fördelar respektive nackdelar med källorna beslutades det att ULF 2002 skulle användas som bas för evalveringsundersökningen. Resultat och rekommendationer, inför evalveringen av UREG, utifrån arbetet med förstudien går att läsa om i "Evalvering av utbildningsregistret – en förstudie, 2003-02-11".

2.1.1 Svensk utbildningsnomenklatur (SUN 2000)

Svensk utbildningsnomenklatur är en standard för klassificering av enskilda utbildningar, samtidigt som den utgör ett system för aggregering av utbildningar till större grupper. SUN används i såväl centrala register för officiell statistik som för registrering av personalens utbildning i personaladministrativa informationssystem.

SUN genomgick under 1998-1999 en välbehövlig revidering. Ett av syftena var att anpassa standarden till den internationella standarden för utbildningsklassificering, ISCED 97. Den nya versionen av SUN – SUN 2000 – gäller som nationell standard för utbildningsklassificering fr.o.m. juli år 2000.

SUN 2000 är uppbyggd av två moduler; en nivåmodul på tre positioner och en inriktningsmodul med fyra positioner. Båda är hierarkiskt uppbyggda vilket möjliggör en flexibel användning genom att samtliga positioner inte behöver användas.

Utbildningsnivå

I nivåmodulen motsvarar den första siffran nivå i ISCED 97, den andra siffran utbildningens längd och den tredje typ av utbildning. I evalveringen av UREG görs jämförelser för en aggregerad nivå av de första två positionerna enligt följande;

00	Utbildning på förskolenivå
10+20	Förgymnasial utbildning
31+32	Gymnasial utbildning 2 år eller kortare
33	Gymnasial utbildning 3 år eller längre
41+52	Eftergymnasial utbildning 2 år eller kortare
53-55	Eftergymnasial utbildning 3 år eller längre
60+62+64	Forskarutbildning t.ex. licentiat och doktor

Utbildningsinriktning

Inriktningsmodulens två första siffror motsvarar inriktning i ISCED 97, den tredje siffran ger också internationell jämförbarhet med "Fields of Education and Training" - ett komplement till ISCED 97. Den fjärde positionen ger en ytterligare specificering. I evalveringen av UREG görs jämförelser enbart för den första positionen i utbildningsinriktning enligt följande;

0	Allmän utbildning
1	Pedagogik och lärarutbildning
2	Humaniora och konst
3	Samhällsvetenskap, juridik, handel, administration
4	Naturvetenskap, matematik och data
5	Teknik och tillverkning
6	Lant- och skogsbruk samt djursjukvård
7	Hälso- och sjukvård samt social omsorg
8	Tjänster
9	Okänd

Klassificering av högsta utbildning

Även om SUN primärt är ett system för att klassificera utbildningar används den också, som i UREG, till att klassificera en individs högsta formella utbildning.

2.2 Syfte

I UREG finns vissa kända felkällor, exempelvis saknas utbildningsnivå för sju procent respektive utbildningsinriktning för tolv procent, av personerna som är födda utomlands. Utbildningsformer inom det reguljära utbildningsväsendet t.ex. grundskola, gymnasium, högskola/universitet och kommunal vuxenutbildning fångas upp av UREG i hög grad, medan kortare yrkesutbildningar, personalutbildning och andra liknande utbildningar knappt täcks in alls. När det gäller personer födda utomlands finns en strävan att få bättre kvalitet på statistiken, medan det är svårare att förbättra kvaliteten för kortare yrkesutbildningar. Eftersom statistiken över befolkningens utbildningsnivå bygger på UREG är det dock viktigt att veta vilka felkällor som finns även om de är svåra att åtgärda.

Syftet med en evalvering av UREG är att mäta tillförlitligheten i statistiken och identifiera eventuella fel som har uppkommit i samband med insamling och bearbetning av uppgifter till registret. Statistikproducenten är dessutom intresserad av att analysera hur felet uppkommer för att därigenom kunna undvika dem i kommande statistik. Kartläggningen kan leda till förbättringar av statistiken.

För att uppfylla huvudmålet för evalveringen delas studien upp i delmål:

- Dokumentera arbetsgången och de metoder som har använts vid evalveringen.
- Beskriva tillförlitligheten i statistiken inom olika delpopulationer t.ex. efter källa, kön, ålder, utbildningsnivå och utbildningsinriktning.
- Identifiera orsaker till eventuella brister i UREG.

Evalveringen avser uppgifterna om högsta genomförda utbildning för individerna i UREG i form av utbildningsnivå på aggregerad två-siffernivå och utbildningsinriktning på en-siffernivå enligt SUN 2000.

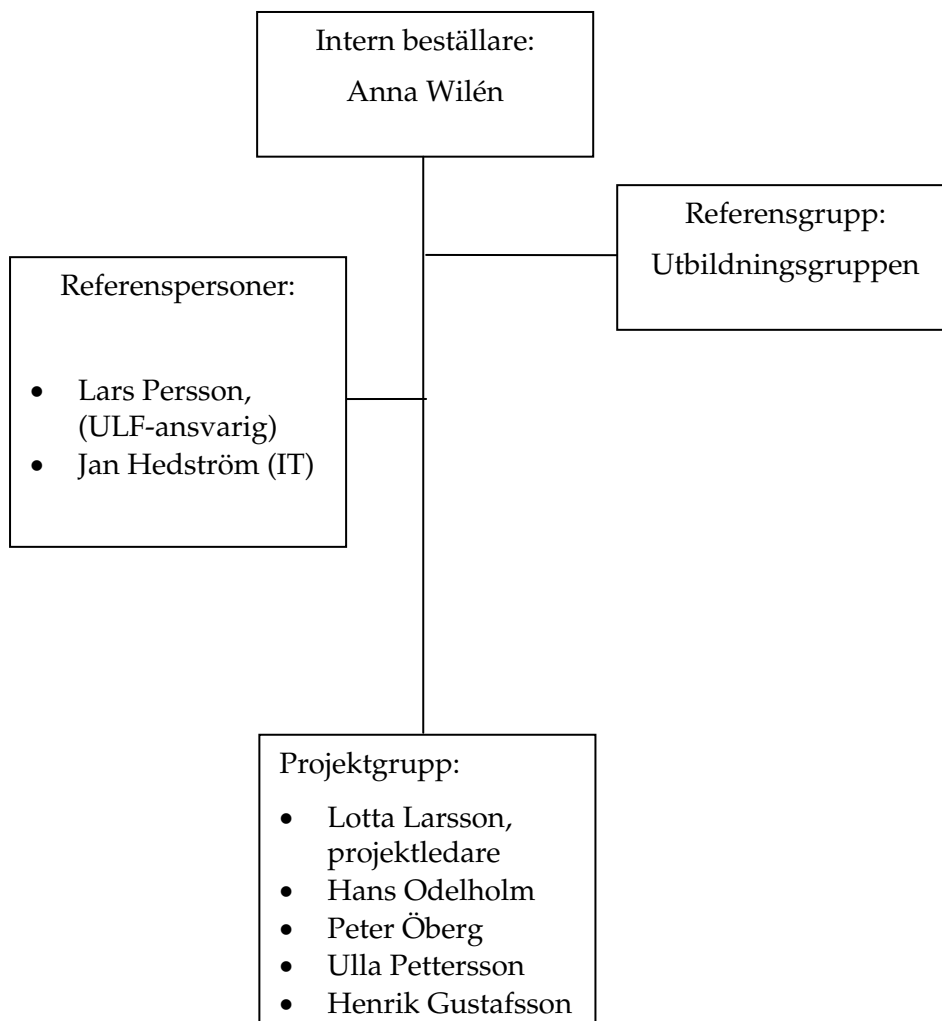
Evalveringen berör endast tillförlitligheten i statistiken över befolkningens utbildning. Skattningarna av fel täcker förmodligen bortfalls-, mät- och bearbetningsfel, medan storleken på eventuella täckningsfel inte berörs av studien. Statistiken över befolkningens utbildning är totalräknad, vilket innebär att urvalsfel inte förekommer. Evalveringen täcker heller inte in aktualitet, innehåll, förståelighet och jämförbarhet som behövs för att ge en helhetsbild av kvaliteten i statistiken från UREG.

I projektet ingår inte att analysera eventuella orsaker till brister i UREG. Detta innebär att tänkbara orsaker till brister kommer att identifieras utan att någon direkt analys av hur allvarliga orsakerna är genomförs. Analys av orsaker till brister i statistiken och förslag till åtgärder för att förbättra tillförlitligheten i statistiken från UREG får genomföras under produktionen av kommande versioner av UREG.

I projektet ingår inte heller att undersöka tillförlitligheten i statistiken från ULF avseende högsta utbildning.

3 Deltagare i projektet

Deltagarna i projektet visas av nedanstående bild. I projektgruppen ingick personal som arbetar med UREG och har ämneskunskap, en expert gällande SUN 2000 och personal med metodstatistisk kompetens.



4 Behov

4.1 Användare

De viktigaste användarna av statistiken över befolkningens utbildning är;

- Utbildningsdepartementet
- Skolverket
- Högskoleverket
- SCB:s prognosinstitut
- Övriga myndigheter inom utbildningsområdet t.ex. IFAU och ITPS
- CSN
- Arbetsmarknadens organisationer
- Länsstyrelser och kommuner
- Arbetsmarknadsstyrelsen (AMS)

Mått på eventuella brister i UREG är av intresse både för användare och producenter. Dessa mått ger en bild av tillförlitligheten i statistiken över befolkningens utbildning samtidigt som de är viktiga för forskare och andra som använder utbildningsnivå som bakgrundsvariabel i analyser av olika slag. Denna form av mått behövs för att skapa förtroende för statistiken. Det är framförallt forskarna som har efterfrågat de mått på statistikens tillförlitlighet som beräknades vid evalveringen 1990.

4.2 Användningsområden

Statistiken över befolkningens utbildning används som underlag för utbildningsplanering både på nationell och regional nivå och statistik över befolkningens utbildning i olika regioner behövs bl. a. för en effektiv fördelning av samhällets utbildningsresurser. Statistiken används också för internationella jämförelser och vid internationell rapportering. Årlig samarbetsbearbetning med SCB:s registerbaserade arbetsmarknadsstatistik (RAMS) ger information om utbildningssituationen inom olika delar av arbetsmarknaden. Statistiken över befolkningens utbildning har dessutom ett stort allmäninformativt värde.

Eftersom utbildningsnivå är en viktig socioekonomisk variabel används UREG också mycket frekvent i olika forskningsprojekt. Exempel är forskning inom arbetsmarknad, integration och medicin.

5 Genomförande

Designen av evalveringen gällande UREG version 2003-01-01 har baserats på hur tidigare evalveringar har gått till dvs. evalveringarna av FoB 90 och UREG 1991. Dessa finns beskrivna i förstudien till evalveringen av UREG (Evalvering av UREG – en förstudie 2003-02-11).

Arbetet med evalveringen av UREG har delats upp i tre faser;

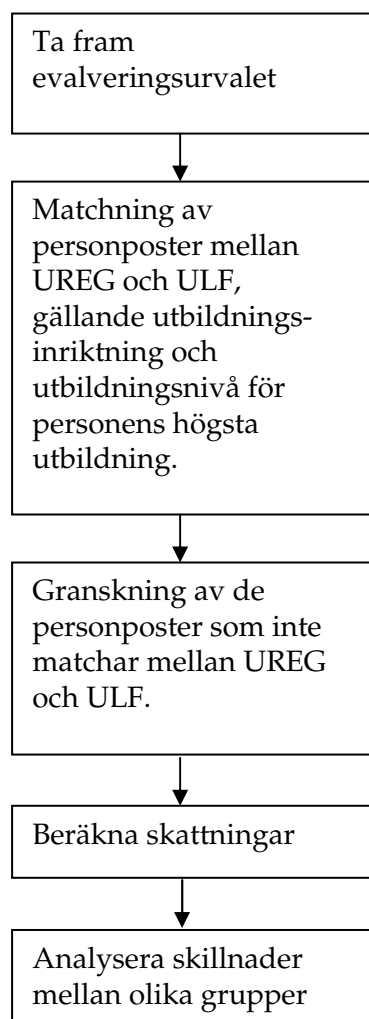
- 1) Ta fram evalveringsurvalet
- 2) Granskning
- 3) Beräkna skattningar gällande fel i uppgifterna om utbildningsnivå respektive utbildningsinriktning i UREG

I första fasen av arbetet skapades urvalet för evalveringen genom att selektera personer i åldern 16 till 74 år ur 3:e respektive 4:e kvartalet av ULF. Anledningen till detta är att åldersavgränsningen för evalveringsstudien bör vara samma som för UREG. Personerna i evalveringsurvalet matchades sedan mot UREG.

I detta arbete ingick också att bestämma vilka variabler ur ULF respektive UREG som behövdes för evalveringen. För samtliga personer i urvalet genomfördes en matchning av utbildningsnivå och utbildningsinriktning för personens högsta utbildning. I de fall som uppgifterna var lika antogs de vara "sanna".

De poster som inte matchade på utbildningsnivå eller utbildningsinriktning gick vidare till granskningen. Syftet med granskningen var i första hand att bestämma en "sann" kod för utbildningsnivå och/eller utbildningsinriktning baserad på tillgänglig information från ULF respektive UREG. I de fall det inte gick att bestämma en "sann" kod sattes "ej kodningsbar". Det bestämdes också en orsakskod till varför UREG och ULF inte matchade innan granskning för utbildningsnivå/utbildningsinriktning.

I den tredje fasen av arbetet beräknades skattningar av andel rätt klassificerade, nettofel och bruttofel för utbildningsnivå/utbildningsinriktning i UREG efter intressanta redovisningsgrupper. Även skattningar avseende olika orsaker till skillnader mellan uppgifter från ULF och UREG togs fram.



5.1 Bestämna "sanna" koder

5.1.1 Tidigare evalveringar

I evalveringen av FoB 90 matchades uppgifterna från AKU respektive FoB 90 mot varandra. I de fall som båda var lika ansågs uppgiften vara "sann". För de fall där uppgifterna inte matchade granskades uppgifterna vid SCB. Syftet med granskningen var att bestämma vilken uppgift som kunde anses vara "sann".

Om det inte gick att bestämma "sann" uppgift gjordes en återintervju med uppgiftslämnaren. Med hjälp av dessa svar bestämdes "sanna" uppgifter.

De "sanna" uppgifterna jämfördes sedan mot urvalets slutgiltiga FoB-uppgifter. Genom att jämföra uppgifterna från kontrollundersökningen med motsvarande uppgifter från den totalräknade Folk- och Bostadsräkningen har man alltså försökt att uppskatta felen i totalräkningen.

I evalveringen av UREG version 1991-01-01 användes det "sanna" registret som togs fram för evalveringen av FoB 90. Även här antogs uppgifterna vara "sanna" om koderna för utbildningsnivå respektive utbildningsinriktning var lika i UREG och det "sanna" registret.

Nedan följer några resultat från evalveringen av UREG version 1991-01-01. Andelen rätt klassificerade avseende utbildningsnivå var 83 procent. Lägst andel som klassificerades rätt fanns i utbildningsnivå fyra (gymnasial utbildning tre år och längre), där andelen var 76,9 procent. Högst andel rätt klassificerade fanns inom eftergymnasial utbildning längre än två år samt forskarutbildning där andelarna var 92,7 procent respektive 93,9 procent. När det gäller utbildningens inriktning var andelen rätt klassificerade 86,1 procent.

I evalveringen av UREG version 2003-01-01 har i stort sett samma procedur, för att bestämma "sanna" uppgifter, använts som i evalveringen av FoB 90. I evalveringen av FoB 90 genomfördes dock återintervjuer med uppgiftslämnare som inte hade någon "sann" uppgift efter granskningen. Detta var av resursskäl inte möjligt i evalveringen av UREG version 2003-01-01. Dessutom jämfördes "sanna" koder mot en slutgiltig version av FoB 90, vilket inte har gjorts i evalveringen av UREG.

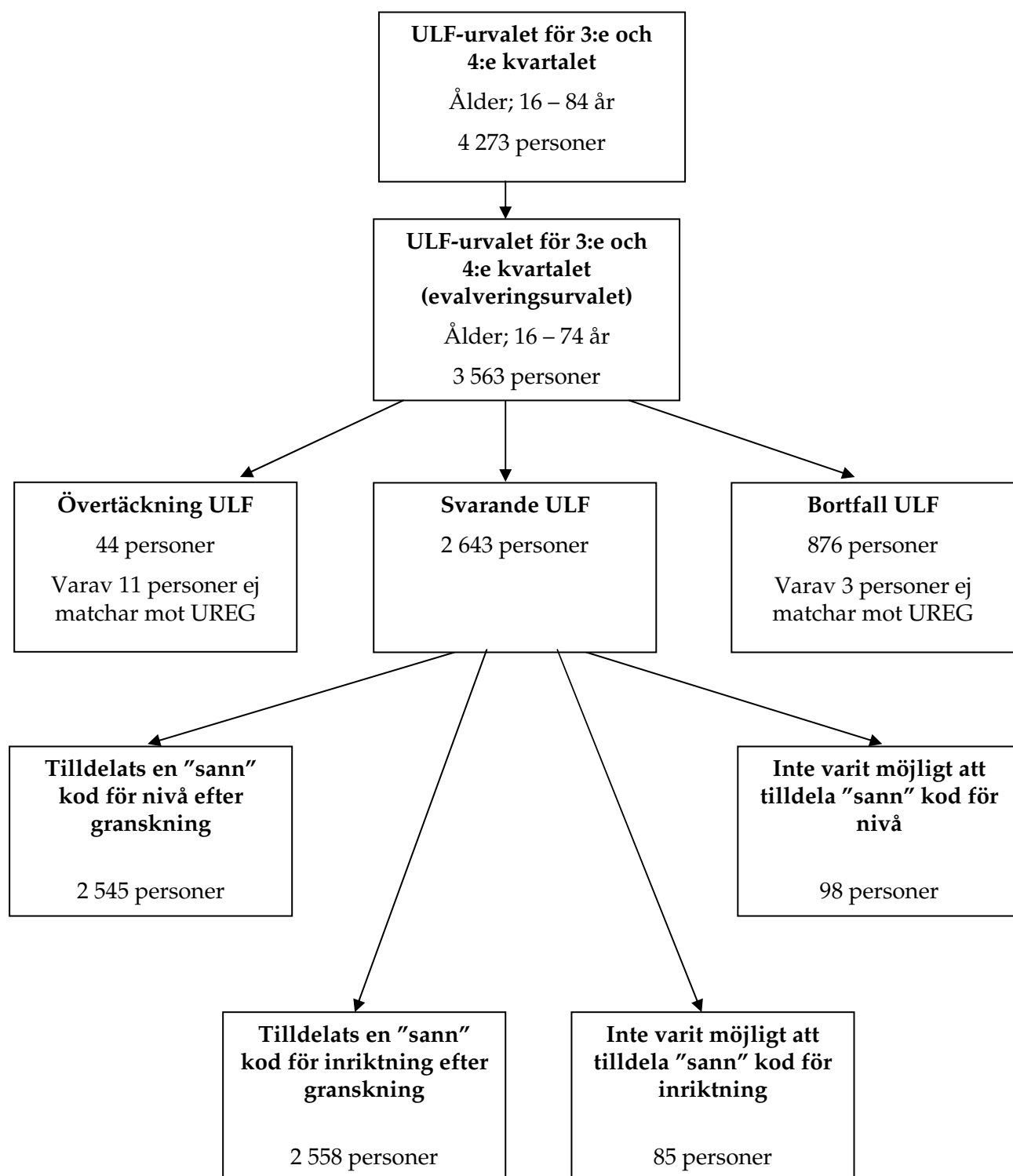
5.1.2 Skapa evalveringsundersökningen

Målpopulationen för UREG består av personer i åldern 16 till 74 år vid tidpunkten 2003-01-01 som stadigvarande är bosatta i Sverige. UREG innehåller 6 423 996 personer. Den undersökning som används för evalveringen av UREG bör täcka målpopulationen för UREG. Målpopulationen för ULF består av personer som fyller 16 – 84 år under 2002 och stadigvarande är bosatta i Sverige. Detta innebär att målpopulationen för evalveringsundersökningen i stort är en delpopulation av ULF. Skillnaden består av att samtliga personer som har invandrat under 2002 inte täcks av ULF p.g.a. urvalsförfarandet. Dessa utgör således undertäckning i evalveringsundersökningen.

Evalveringsurvalet tas fram genom att en selektering görs från ULF med avseende på ålder. De personer som inte matchar mellan evalveringsurvalet och UREG antas ha utvandrat eller avlidit och därmed utgör de övertäckning i evalveringsundersökningen. Det finns också bortfall i ULF vilket således också blir bortfall i evalveringsundersökningen.

För de svarande i evalveringsundersökningen har det inte alltid gått att bestämma en "sann" uppgift för utbildningsnivå respektive utbildningsinriktning. De "ej kodningsbara" utgör bortfall i evalveringsundersökningen vid beräkningar av skattningar gällande fel i uppgifterna från UREG.

Evalveringsundersökningen innehåller både variabler från ULF och UREG, vilka har hämtats via en matchning mellan ULF-urvalet och UREG. Samtliga variabler som behövs för granskningen och vid analysen av tillförlitligheten i statistiken ingår i evalveringsundersökningen. Under granskningen skapades några nya variabler. Exempel på variabler som har skapats via matchning och granskning är "sann" uppgift för utbildningsnivå respektive utbildningsinriktning. På nästa sida visas antal objekt i urval, svars-mängd och bortfall för evalveringsundersökningen.



5.1.3 Matchning mellan UREG och ULF

Det första steget för att fastställa "sann" kod för utbildningsnivå respektive utbildningsinriktning var att genomföra en matchning av personer i UREG mot svarande i ULF. Matchningen avsåg uppgifterna för personernas högsta utbildning. I de fall uppgifterna för utbildningsnivå respektive utbildningsinriktning var lika i UREG och ULF antogs de vara "sanna".

De poster som inte matchade på utbildningsnivå eller utbildningsinriktning gick vidare till granskningen. Det var 796 personer som gick vidare till granskningsproceduren, som utgjorde det andra steget för att fastställa "sanna" koder. Nedan finns några tabeller med resultat från matchningen innan granskning.

**Andel personer med lika kod för utbildningsnivå i UREG och ULF innan granskning med fördelning efter utbildningsnivå enligt UREG.
Andel personer av svarande i ULF**

Utbildningsnivå	Antal svar	Andel matchande
Förgymnasial utbildning	630	73
Gymnasial utbildning 2 år och kortare	721	74
Gymnasial utbildning 3 år	474	72
Eftergymnasial utbildning 2 år och kortare	356	81
Eftergymnasial utbildning 3 år och längre	419	95
Forskarutbildning	25	100
Okänd utbildningsnivå	18	0
Totalt	2643	77

**Andel personer med lika kod för utbildningsinriktning i UREG och ULF innan granskning med fördelning efter utbildningsinriktning enligt UREG.
Andel personer av svarande i ULF**

Utbildningsinriktning	Antal svar	Andel matchande
Allmän utbildning	825	71
Pedagogik och lärarutbildning	159	84
Humaniora och konst	85	84
Samhällsvetenskap, juridik, handel och administration	434	82
Naturvetenskap, matematik och data	60	75
Teknik och tillverkning	494	85
Lant- och skogsbruk samt djursjukvård	52	90
Hälso- och sjukvård samt social omsorg	341	88
Tjänster	143	69
Okänd utbildningsinriktning	50	0
Totalt	2643	78

5.1.4 Granskning

Processbeskrivning

Alla poster har granskats i grupp av personal med ämneskompetens och metodkompetens. Varje post har diskuterats noga innan den "sanna" koden har bestämts.

De poster som hade olika värden i ULF och UREG (nästan 800 personer) gick vidare till en manuell granskning. Syftet med granskningen var att fastställa en "sann" SUN-kod genom att använda hjälpinformation i ULF och UREG för respektive individ.

Den hjälpinformation som användes från ULF-materialet var uppgifter i klartext om genomförd utbildning inom gymnasieskolan, högskola/ universitet, komvux samt övrig utbildning.

Utöver denna information var även uppgifter om kön, ålder, sysselsättningsstatus, yrke och studiedeltagande under läsåret 2001/02 registrerade.

Hjälpinformationen från UREG bestod av SUN 2000-kod, källår, examens-termin och examenskommun för de tre högsta utbildningarna som fanns i utbildningsregistrets bruttofil för respektive individ.

Ett särskilt granskningsformulär användes vid granskningen av de icke matchande posterna, där ovan nämnda hjälpinformation visades för varje individ.

Utifrån ULF-data och data från UREG bestämdes för varje post genom en samlad bedömning om någon kod var sann/rimlig och i så fall vilken. I annat fall bedömdes posten som ej kodningsbar.

Kriterier som har använts för att fastställa "sann" kod

I resultaten redovisas UREGs uppgiftskällor aggregerade till fyra grupper:

Reguljära statistikregister är huvudsakligen register över avgångna och examinerade från de reguljära utbildningsformerna. FoB 90 är ett antal olika varianter av Folk- och bostadsräkningen 90. Övriga administrativa register består bland annat av databas för komvux-poäng, arbetsmarknadsutbildningar och en enkätundersökning om utbildning utanför Sverige. Källor med låg prioritet är källor där personers utbildningsbakgrund är registrerad, huvudsakligen utanför SCB. Exempel är centrala studiestödsnämnden, svenska för invandrare och arbetsmarknadsstyrelsens register över arbetssökande.

Vid granskningen och fastställandet av "sann" utbildning togs viss hänsyn till kvaliteten på utbildningsregistrets källor. Grundprincipen var dock att uppgiftslämnarna hade angett riktiga uppgifter i ULF, om inte uppgifterna var ofullständiga eller det fanns annan information som talade emot att uppgifterna var riktiga.

Andel matchande efter granskning

Efter granskningen var det fler personer vars uppgifter var lika i UREG och evalveringsundersökningen än vad det var innan granskningen. Effekten av granskningen kan utläsas ur tabellerna nedan.

**Andel personer med lika kod för utbildningsnivå i UREG och evalveringsstudien efter granskning med fördelning efter utbildningsnivå enligt UREG.
Andel personer av svarande i evalveringsstudien**

Utbildningsnivå	Antal svar	Andel matchande
Förgymnasial utbildning	619	78
Gymnasial utbildning 2 år och kortare	685	89
Gymnasial utbildning 3 år	448	85
Eftergymnasial utbildning 2 år och kortare	334	89
Eftergymnasial utbildning 3 år och längre	416	99
Forskarutbildning	25	100
Okänd utbildningsnivå	18	0
Totalt	2545	86

**Andel personer med lika kod för utbildningsinriktning i UREG och evalveringsundersökningen efter granskning med fördelning efter utbildningsinriktning enligt UREG.
Andel personer av svarande i evalveringsstudien**

Utbildningsinriktning	Antal svar	Andel matchande
Allmän utbildning	796	78
Pedagogik och lärarutbildning	151	95
Humaniora och konst	81	96
Samhällsvetenskap, juridik, handel och administration	417	94
Naturvetenskap, matematik och data	59	90
Teknik och tillverkning	482	96
Lant- och skogsbruk samt djursjukvård	51	98
Hälso- och sjukvård samt social omsorg	334	93
Tjänster	137	90
Okänd utbildningsinriktning	50	0
Totalt	2558	87

5.2 Metoder för att ta fram skattningar

5.2.1 Mått på tillförlitligheten i statistiken

I tidigare evalveringsundersökningar har antal rätt klassificerade, bruttofel samt nettofel tagits fram.

Vilket mått som är av störst intresse för användare av statistiken från UREG beror på hur statistiken används. För dem som endast använder statistik från UREG på den nivå som redovisas i tabeller är nettofelen mest intressanta. För att bedöma tillförlitligheten i statistik för mindre redovisningsgrupper bör man även titta på antal rätt klassificerade och bruttofel. Dessa kvalitetsmått är även av intresse då UREG används för stratifiering vid urvals-

dragning, som hjälpinformation vid estimation, variansanalyser, registerutdrag etc.

Metoden för att bestämma "sanna" koder påverkar storleken på skattningarna av fel i hög grad. I denna evalvering har inga återkontakter gjorts för att bestämma "sanna" koder för utbildning. Detta innebär att det inte är den exakta storleken på skattningarna av fel som är av intresse, utan snarare om det finns skillnader mellan olika grupper av populationen eller om det är någon viss utbildningsform som orsakar problem. Skattningarna av fel kan även tolkas som skattningar av skillnader mellan UREG och evalveringsundersökningen. I de fall som det finns stora skillnader finns det förmodligen också problem med uppgifterna om högsta utbildning i UREG.

För producenten av statistiken är det också av intresse att analysera bakomliggande orsaker till felen för att på så sätt kunna åtgärda en del av felen till kommande undersökningar. I evalveringen av UREG version 2003-01-01 är vi därför också intresserade av orsaker till att det har angivits olika utbildningsnivå och/eller utbildningsinriktning i UREG och ULF. Nedan visas två tabeller med olika parametrar av intresse för evalveringen av UREG.

Parametrar av intresse gällande orsaker till skillnader mellan UREG och ULF

Orsaker till skillnader i utbildningsnivå eller utbildningsinriktning bestämdes för alla personer, där utbildningsnivå/utbildningsinriktning inte matchade mellan ULF och UREG innan granskning. De parametrar som är av intresse för att beskriva orsaker till skillnader mellan UREG och ULF visas i nedanstående tablåer.

Innan granskning

Svarande i ULF	UREG version 2003-01-01			Totalt
	Grupp 1	Grupp 2	Uppgift saknas	
Redovisningsgrupp 1	N_{11}	N_{12}	N_{1B}	$N_{1.}$
Redovisningsgrupp 2	N_{21}	N_{22}	N_{2B}	$N_{2.}$
Uppgift saknas	N_{B1}	N_{B2}	..	$N_{B.}$
Antal enligt UREG	$N_{.1}$	$N_{.2}$	$N_{.B}$	$N_{..}$
Antal enligt ULF	$N_{.1}$	$N_{.2}$	$N_{.B}$	$N_{..}$
Antal matchande	N_{11}	N_{22}	..	$N_{11} + N_{22}$
Antal ej matchande	$N_{.1} - N_{11}$	$N_{.2} - N_{22}$	$N_{.B}$	$N_{..} - N_{11} - N_{22}$

De poster som inte matchade mellan UREG och ULF, t.ex. Antejmatch=N..-N11- N22-...-N66 för utbildningsnivå, sprids över orsakskoder. Ett exempel för utbildningsnivå visas nedan.

Utbildningsnivå enligt		Orsaker till skillnader					Totalt
UREG	ULF	1	2	3	4	Ingen "sann" kod	
Förgym, i=1	Ej förgym, j<>1	N _{1j1}	N _{1j2}	N _{1j3}	N _{1j4}	N _{1j5}	N _{.1} - N ₁₁
Kort gym, i=2	Ej kort gym, j<>2	N _{2j1}	N _{2j2}	N _{2j3}	N _{2j4}	N _{2j5}	N _{.2} - N ₂₂
Lång gym, i=3	Ej lång gym, j<>3	N _{3j1}	N _{3j2}	N _{3j3}	N _{3j4}	N _{3j5}	N _{.3} - N ₃₃
Kort egym, i=4	Ej kort egym, j<>4	N _{4j1}	N _{4j2}	N _{4j3}	N _{4j4}	N _{4j5}	N _{.4} - N ₄₄
Lång egym, i=5	Ej lång egym, j<>5	N _{5j1}	N _{5j2}	N _{5j3}	N _{5j4}	N _{5j5}	N _{.5} - N ₅₅
Forskutb, i=6	Ej forskutb, j<>6	N _{6j1}	N _{6j2}	N _{6j3}	N _{6j4}	N _{6j5}	N _{.6} - N ₆₆
Okänd utb, i=9	Ej okänd utb, j<>9	N _{9j1}	N _{9j2}	N _{9j3}	N _{9j4}	N _{9j5}	N _{.B}
Totalt, i<>j	Totalt, i<>j	N _{i<>j1}	N _{i<>j2}	N _{i<>j3}	N _{i<>j4}	N _{i<>j5}	Antejmatch

Av de matchande som det går att bestämma "sann" kod för skattas andelen personer efter olika orsaker till skillnader mellan UREG och ULF enligt följande exempel;

$$\text{Andel med orsak}=1 \text{ för förgymnasial utbildning} = \frac{N_{1j1}}{N_{.1} - N_{11} - N_{1j5}}$$

Parametrar av intresse gällande fel i UREG

De parametrar som är av intresse att skatta utifrån evalveringsstudien är andel rätt klassificerade, bruttofel och nettofel. Antal rätt klassificerade definieras som antalet riktigt medtagna inom en viss utbildningsnivå/utbildningsinriktning enligt UREG. Bruttofelet är definierat som summan av antalet felaktigt uteslutna och antalet felaktigt medtagna. Nettofelet definieras som skillnaden mellan skattningen av antalet personer enligt UREG och antalet personer enligt evalveringsundersökningen, inom en viss utbildningsnivå/utbildningsinriktning. Detta innebär att de individuella felen kan ta ut varandra genom att antalet felaktigt medtagna är ungefär lika stort som antalet felaktigt uteslutna. Nedan följer formeluttryck för dessa parametrar.

		Utbildningsregistret version 2003-01-01			
Tilldelats "sann" utbildningskod Evalveringsundersökning (EU)		Redovisnings- grupp 1	Redovisnings- grupp 2	Uppgift saknas	Totalt
	Redovisningsgrupp 1	N_{11}	N_{12}	N_{1B}	$N_{1.}$
	Redovisningsgrupp 2	N_{21}	N_{22}	N_{2B}	$N_{2.}$
A	Antal enl. UREG	$N_{.1}$	$N_{.2}$	$N_{.B}$	$N_{..}$
B	Antal enl. EU	$N_{1.}$	$N_{2.}$..	$N_{1.}+N_{2.}$
C	Antal rätt klassific.	N_{11}	N_{22}	..	$N_{11}+N_{22}$
D	Andel rätt klass. %	$100*N_{11}/N_{1.}$	$100*N_{22}/N_{2.}$..	$100*(N_{11}+N_{22})/N_{..}$
E	Felaktigt medtagna i redovisningsgruppen	$N_{.1}-N_{11}$	$N_{.2}-N_{22}$	$N_{.B}$	$N_{..}-N_{11}-N_{22}$
F	Felaktigt uteslutna i redovisningsgruppen	$N_{1.}-N_{11}$	$N_{2.}-N_{22}$		$N_{1.}+N_{2.}-N_{11}-N_{22}$
G	Bruttofel	$N_{.1}+N_{1.}-2N_{11}$	$N_{.2}+N_{2.}-2N_{22}$	$N_{.B}$	$N_{..}+N_{1.}+N_{2.}-2N_{11}-2N_{22}$
H	Nettofel	$N_{.1}-N_{1.}$	$N_{.2}-N_{2.}$		$N_{..}-N_{1.}-N_{2.}$
I	Relativt nettofel i procent	$100*(N_{.1}-N_{1.})/N_{.1}$	$100*(N_{.2}-N_{2.})/N_{.2}$		$100*(1-(N_{1.}+N_{2.})/N_{..})$

Där;

N_j = Antal personer

i = Redovisningsgrupp i evalveringsundersökningen (1...i...I) t.ex. 1 avser förgymnasial utbildning etc.
(B = uppgift saknas).

j = Redovisningsgrupp i Utbildningsregistret (1...j...J) t.ex. 1 avser förgymnasial utbildning etc.
(B = uppgift saknas).

5.2.2 Urvalsdesign

Undersökningarna av levnadsförhållanden

Urvalet i ULF består av individer 16-84 år och dras från registret över totalbefolkningen, RTB. Det årliga urvalet om ca. 7 500 personer består i själva verket av flera delurval. Ett systematiskt urval är draget inom varje delpopulation.¹

Vid estimationen antas urvalsdesignen vara OSU i ULF. Skattningarna i ULF bygger på ett poststratifieringsförfarande. Urvalet och populationen delas upp i 96 grupper (poststrata) efter kombinationer av kön, åldersklasser, H-region, civilstånd och fått lön under året². Det antas att inom varje grupp är svarssannolikheten lika och individerna svarar oberoende av varandra.

¹ Se studie av bortfallet i 2000 års undersökning om levnadsförhållanden (Hassan Mirza).

² För en detaljerad beskrivning av poststrata se Levnadsförhållanden, Appendix 15 (1995).

I en studie av bortfallet avseende 2000 års ULF³ visade det sig att bortfallet snedvred skattningar för några registervariabler som användes i studien. Antalet personer födda i Sverige, personer med relativt hög utbildning samt sysselsatta personer överskattades medan antalet personer födda utomlands, personer med relativt låg utbildning samt personer i arbetsmarknadspolitiska åtgärder underskattades.

Evalveringsstudien

I evalveringsstudien har det antagits att urvalet har dragits med stratifierat obundet slumpmässigt urval vid urvalstillfället (STOSU). De 96 grupper (poststrata) som används när skattningarna från ULF tas fram, antas dela upp urvalsramen i grupper (strata) innan urvalsdragningen har genomförts. Inom varje stratum antas sedan ett obundet slumpmässigt urval ha dragits. När skattningarna från ULF tas fram genomförs i själva verket indelningen i grupper vid estimationstillfället och inte vid urvalstillfället.

5.2.3 Skattningsförfarande

Hjälpinformation från UREG version 2003-01-01 används vid estimationen. Detta estimationsförfarande, som utnyttjar hjälpinformation i form av registervariabler, används i det här fallet främst för att kompensera för den skevhet som kan antas uppstå till följd av bortfall. Förfarandet kan också reducera den slumpmässiga osäkerheten (variansen) som skattningarna är behäftade med.

Felen i utbildningsinriktning respektive utbildningsnivå skattas var för sig med olika uppsättningar av hjälpvariabler, vilket innebär att två olika kalibreringsestimatorer används. Hjälpvektorerna för skattningar av fel för utbildningsnivå/ utbildningsinriktning bildas enligt följande;

- För utbildningsnivå används;
 $Utbildningsinriktning + Utbildningsnivå + Källa*nivå + Åldersgrupp*nivå +$
 $Kön*nivå + Född i Sverige/Född utomlands*nivå$
- För utbildningsinriktning används;
 $Utbildningsinriktning + Utbildningsnivå + Källa*inriktning +$
 $Åldersgrupp*inriktning + Kön* inriktning +$
 $Född i Sverige/Född utomlands*inriktning$

En mer detaljerad beskrivning av skattningsförfarandet som har använts i evalveringsstudien finns i bilaga 1.

³ Bakgrundsfakta 2003:8, Hassan Mirza.

6 Resultat

6.1 Tillförlitligheten i statistiken

Enligt evalveringsstudien var punktskattningen av andelen rätt klassificerade 85 procent för utbildningsnivå i UREG.

Skattningar av fel i UREG med avseende på utbildningsnivå

Utbildningsnivå	Antal personer enligt;		Andel av personer i UREG		
	UREG	"sann" kod	Nettofel	Rätt klassificerade	Bruttofel
Förgymnasial utbildning	1 663 300	1 348 100	19 ± 4	77 ± 4	459 400 ± 59 100
Gymnasial utbildning 2 år och kortare	1 728 100	1 799 300	-4 ± 4	88 ± 3	476 400 ± 65 700
Gymnasial utbildning 3 år	1 173 500	1 231 000	-5 ± 6	84 ± 4	443 900 ± 65 100
Eftergymnasial utbildning 2 år och kortare	819 100	943 800	-15 ± 7	89 ± 4	310 100 ± 53 700
Eftergymnasial utbildning 3 år och längre	891 600	1 048 100	-18 ± 5	99 ± 2	180 100 ± 43 800
Forskarutbildning	49 700	53 700	-8 ± 12	100	4 100 ± 5 700
Okänd utbildning	98 700	0	100	0	98 700
Totalt	6 424 000	6 424 000		85 ± 2	1 972 700 ± 173 500

Skattningarna av nettofelet tyder på att antalet personer med förgymnasial utbildning överskattas i UREG medan antalet personer inom samtliga övriga utbildningsnivåer underskattas. Detta mönster fanns även i den tidigare evalveringen av UREG 1991. Andelen rätt klassificerade med avseende på utbildningsnivå var marginellt högre i denna evalvering jämfört med evalveringen av UREG 1991. Det är dock svårt att jämföra resultaten för dessa evalveringar eftersom det finns skillnader i metoder t.ex. har ULF använts i denna evalvering medan AKU användes för evalveringen 1991. Dessutom har ingen telefonuppföljning genomförts vid bestämningen av "sanna" värden.

Skattningarna av nettofel visar att underskattningen är högre för eftergymnasial utbildning än för gymnasial utbildning. Underskattningen av antalet personer med gymnasial utbildning i UREG är inte statistiskt säkerställd. Skattningarna av Bruttofel visar att summan av olika fel är stora för både gymnasiala utbildningar och eftergymnasiala utbildningar. Detta tyder på att felaktigt uteslutna och felaktigt medtagna tar ut varandra för gymnasiala utbildningar.

De största problemen i UREG, med tanke på skattningarna av fel, tycks finnas vid klassificeringen av gymnasial utbildning tre år respektive eftergymnasial utbildning två år eller kortare.

För forskarutbildning ska man vara medveten om att skattningarna av fel i UREG är relativt osäkra, vilket gör att resultaten är svåra att tolka.

Skattad andel personer av befolkningen efter utbildningsnivå enligt UREG och utbildningsnivå enligt evalveringsundersökningen

UREG	Evalveringsundersökning						Tot
	För Gym	Gym <=2 år	Gym 3 år	Egym <3 år	Egym >=3 år	Forsk Utb	
FörGym	77	16	5	1	1	0	100
Gym<=2 år	0	88	9	2	1	0	100
Gym 3 år	0	0	84	13	3	0	100
Egym<3 år	0	0	0	89	11	0	100
Egym>=3 år	0	0	0	1	99	0	100
ForskUtb	0	0	0	0	0	100	100
Okänd utbildning	70	0	5	12	13	0	100

Skattningarna av andel rätt klassificerade som finns i diagonalen i tabellen ovan visar att andelen rätt klassificerade i de flesta fall ökar med utbildningsnivå, vilket överensstämmer med den tidigare evalveringen. När det finns skillnad i utbildningsnivå mellan evalveringsundersökningen och UREG har utbildningsnivån oftast höjts ett steg i evalveringsundersökningen. De felaktigt klassificerade med förgymnasial utbildning i UREG är exempelvis oftast klassificerade inom kort gymnasial utbildning i evalveringsstudien. Personer med okänd utbildningsnivå i UREG bidrar också till felaktiga uppgifter om högsta utbildning i UREG. Majoriteten av personerna med okänd utbildningsnivå i UREG skulle ha varit klassificerade inom förgymnasial utbildning enligt evalveringsstudien.

Enligt evalveringsstudien var punktskattningen av andelen rätt klassificerade i UREG 85 procent för utbildningsinriktning, d.v.s. lika hög som för utbildningsnivå.

Skattningar av fel i UREG med avseende på utbildningsinriktning

Utbildningsinriktning	Antal personer enligt;		Andel av personer i UREG		
	UREG	"sann" kod	Nettofel	Rätt klassificerade	Bruttofel
Allmän utbildning	2 164 000	1 826 900	16 ± 4	77 ± 3	663 400 ± 72 800
Pedagogik och lärarutbildningar	356 700	356 300	0 ± 7	94 ± 5	40 000 ± 21 800
Humaniora och konst	209 000	281 100	-35 ± 15	97 ± 4	86 500 ± 31 100
Samhällsvetenskap, juridik etc.	931 300	1 051 000	-13 ± 6	94 ± 3	231 800 ± 48 800
Naturvetenskap, matematik etc.	121 600	147 400	-21 ± 19	89 ± 9	52 800 ± 23 000
Teknik och tillverkning	1 218 600	1 402 400	-15 ± 5	96 ± 2	292 800 ± 53 400
Lant- och skogsbruk etc.	103 400	120 900	-17 ± 16	98 ± 5	22 200 ± 15 700
Hälso- och sjukvård etc.	736 700	828 000	-12 ± 6	93 ± 3	195 000 ± 43 700
Tjänster	361 200	410 000	-14 ± 11	88 ± 6	137 000 ± 37 700
Okänd utbildning	221 500	0	100	0	221 600
Totalt	6 424 000	6 424 000	0	85±2	1 943 000 ± 159 300

Skattningarna av nettofelen tyder på att antalet personer med allmän utbildning överskattas i UREG medan antalet personer inom samtliga övriga utbildningsinriktningar, utom pedagogik och lärarutbildningar, underskattas. En stor andel av personerna med allmän utbildning har förgymnasial utbildning, vilket gör att det finns ett samband mellan felen i utbildningsnivå och felen i utbildningsinriktning. Andelen rätt klassificerade är således lägst för allmän utbildning.

För de små utbildningsinriktningarna t.ex. "Humaniora och konst", "Naturvetenskap, matematik" och "Tjänster" är skattningarna väldigt osäkra, vilket gör att det är svårt att tolka resultaten. Konfidensintervallen för skattningar av nettofel för "Lant- och skogsbruk" och "Naturvetenskap, matematik" täcker nästan in noll.

Minst problem med tanke på skattningarna av fel tycks finnas vid klassificeringen av pedagogik och lärarutbildningar.

Skattad andel personer av befolkningen efter utbildningsinriktning enligt UREG och utbildningsinriktning enligt evalveringsundersökningen

Utbildningsregistret	Evalveringsundersökning									Totalt
	Allmän	Pedagog	Humaniora	Samhäll.	Naturvet.	Teknik	Lantbruk	Sjukvård	Tjänster	
Allmän	77	0	2	5	1	7	1	5	2	100
Pedagog	0	94	1	1	0	0	0	3	1	100
Humaniora	0	0	97	0	0	0	0	0	3	100
Samhäll.	1	0	1	94	1	1	0	1	1	100
Naturvet.	0	0	0	9	89	0	0	0	2	100
Teknik	1	0	0	1	1	96	0	1	0	100
Lantbruk	0	0	0	0	0	2	98	0	0	100
Sjukvård	2	1	0	2	0	1	0	93	1	100
Tjänster	4	1	0	3	0	3	0	1	88	100
Okända	51	3	6	10	2	21	0	4	3	100

Andelen rätt klassificerade visas i diagonalen av tabellen ovan. Enligt evalveringsstudien tillhör de felaktigt klassificerade i UREG flera olika utbildningsinriktningar. För allmän utbildning tillhör 17 procent samhällsvetenskaplig, teknisk eller vårdutbildning i evalveringsstudien. Felaktigt klassificerade med pedagogisk utbildning i UREG har i flera fall blivit klassificerade som vårdutbildning i evalveringsstudien.

Drygt hälften av personerna med en okänd utbildningsinriktning i UREG har blivit klassificerade som allmän utbildning i evalveringsstudien. Därefter har nästan en tredjedel blivit klassificerade som samhällsvetenskaplig- eller teknisk utbildning.

6.1.1 Skillnader mellan olika grupper i populationen

Inom några grupper av populationen verkar det finnas mer problem med klassificeringen av utbildningsnivå i UREG än i andra grupper.

**Skattningar av fel i UREG med avseende på utbildningsnivå.
Efter källa och utbildningsnivå**

Källa Utbildningsnivå	Antal personer enligt;		Andel av UREG;	
	UREG	"sann" kod	Rätt klass	Nettofel
Lågt prioriterade källor UREG	471 000	471 000	60 ± 7	0
Förgymnasial utbildning	97 100	117 500	50 ± 20	-21 ± 29
Gymnasial utbildning	261 600	305 200	88 ± 7	-17 ± 10
Eftergymnasial utbildning
Okänd utbildningsnivå	98 700	0	0	100
Övriga administrativa register	765 700	765 700	86 ± 5	0
Förgymnasial utbildning
Gymnasial utbildning	619 400	584 500	86 ± 5	6 ± 5
Eftergymnasial utbildning	119 800	163 200	91 ± 9	-36 ± 20
Okänd utbildningsnivå	0	0	-	-
FoB 1990	2 297 300	2 297 300	81 ± 3	0
Förgymnasial utbildning	1 114 000	843 400	76 ± 5	24 ± 5
Gymnasial utbildning	843 700	1 021 400	85 ± 4	-21 ± 7
Eftergymnasial utbildning	339 600	432 500	91 ± 5	-27 ± 9
Okänd utbildningsnivå	0	0	-	-
Reguljära statistikregister	2 890 000	2 890 000	91 ± 2	0
Förgymnasial utbildning	425 800	369 100	87 ± 6	13 ± 6
Gymnasial utbildning	1 176 800	1 119 200	88 ± 4	5 ± 4
Eftergymnasial utbildning	1 287 400	1 401 700	96 ± 2	-9 ± 3
Okänd utbildningsnivå	0	0	-	-

Personer med okänd utbildningsnivå i UREG ingår endast i lågt prioriterade källor till UREG och har fått en "sann" kod, vilken överensstämmer med koden som är angiven i ULF. Eftersom personer med okänd utbildningsnivå i UREG fördelas över "sanna" utbildningsnivåer enligt ULF, blir det underskattningar av både personer med förgymnasial- och gymnasial utbildning för lågt prioriterade källor till UREG.

De största grupperna av källor till UREG är FoB 1990 och reguljära statistikregister. Indelningen efter utbildningsnivå verkar fungera bättre för personer vars uppgifter om högsta utbildningsnivå har hämtats från reguljära statistikregister, jämfört med personer vars uppgifter har hämtats från FoB 1990. I FoB 1990 har uppgiften om högsta utbildningsnivå satts utifrån frågor till befolkningen om högsta utbildningsnivå respektive genomförda utbildningar, dvs. på ungefär samma sätt som i ULF.

Den vanligaste orsaken till att utbildningsnivå inte matchar för källan FoB 90 är att utbildning som anges i ULF saknas i UREG, dvs. personen har inte nämnt utbildningen i FoB 90. Det handlar ofta om utbildningar utanför det reguljära utbildningssystemet, t.ex. längre lärlingsutbildningar eller andra icke nivåspecifiserade utbildningar som exempelvis "Bageri 1 år" och "Snickarlärling 4 år".

Efter ålder och utbildningsnivå

Ålder Utbildningsnivå	Antal personer enligt;		Andel av UREG;	
	UREG	"sann" kod	Rätt klass	Nettofel
16 – 24	937 200	937 200	86 ± 4	0
Förgymnasial utbildning	386 200	403 900	92 ± 5	-5 ± 7
Gymnasial utbildning	374 500	369 700	90 ± 5	1 ± 7
Eftergymnasial utbildning	128 300	163 600	89 ± 9	-28 ± 14
Okänd utbildningsnivå	48 200	0	0	100
25 – 44	2 431 600	2 431 600	84 ± 3	0
Förgymnasial utbildning	294 400	169 700	56 ± 11	42 ± 11
Gymnasial utbildning	1 258 700	1 264 500	85 ± 4	-1 ± 4
Eftergymnasial utbildning	860 600	997 400	95 ± 3	-16 ± 5
Okänd utbildningsnivå	18 000	0	0	100
45 – 64	2 314 200	2 314 200	84 ± 3	0
Förgymnasial utbildning	612 600	469 900	74 ± 7	23 ± 7
Gymnasial utbildning	1 023 900	1 096 100	86 ± 4	-7 ± 5
Eftergymnasial utbildning	654 600	748 200	93 ± 4	-14 ± 5
Okänd utbildningsnivå	23 100	0	0	100
65 – 74	741 000	741 000	88 ± 4	0
Förgymnasial utbildning	370 200	304 600	82 ± 6	18 ± 6
Gymnasial utbildning	244 500	300 000	93 ± 5	-23 ± 10
Eftergymnasial utbildning	117 000	136 400	100	-17 ± 11
Okänd utbildningsnivå	9 400	0	0	100

Andelen personer som har okänd utbildningsnivå i UREG är störst i den yngsta åldersgruppen, fem procent. I övriga åldersgrupper är andelen ungefär en procent. Dessa personer har blivit spridda över utbildningsnivåer enligt den "sanna" koden, vilken är satt utifrån uppgifter i ULF.

I åldersgruppen 16 till 24 år verkar felen vara små för personer med förgymnasial utbildning respektive gymnasial utbildning. I övriga åldersgrupper verkar antalet personer med förgymnasial utbildning vara över-skattade i UREG. För gymnasial utbildning verkar felen även vara små i åldersgruppen 25 – 44 år medan det tycks finnas en underskattning av antalet personer med gymnasial utbildning i åldersgrupperna 45 – 64 år och 65 – 74 år i UREG. Detta hänger troligen samman med att yrkesutbildningar utanför det reguljära utbildningsväsendet, som är relativt vanliga bland de äldre personerna, inte fångas upp via källorna till UREG.

Antalet personer med eftergymnasial utbildning verkar underskattade i samtliga åldersgrupper i UREG.

Efter kön och utbildningsnivå

Kön Utbildningsnivå	Antal personer enligt;		Andel av UREG;	
	UREG	"sann" kod	Rätt klass	Nettofel
Män	3 235 500	3 235 500	84 ± 2	0
Förgymnasial utbildning	868 700	713 800	77 ± 5	18 ± 6
Gymnasial utbildning	1 482 300	1 549 600	85 ± 4	-4 ± 4
Eftergymnasial utbildning	820 800	972 100	95 ± 3	-18 ± 5
Okänd utbildningsnivå	63 800	0	0	100
Kvinnor	3 188 500	3 188 500	86 ± 2	0
Förgymnasial utbildning	794 600	634 300	77 ± 5	20 ± 6
Gymnasial utbildning	1 419 300	1 480 700	88 ± 3	-4 ± 4
Eftergymnasial utbildning	939 700	1 073 500	93 ± 3	-14 ± 4
Okänd utbildningsnivå	34 900	0	0	100

Andelen personer som har okänd utbildningsnivå i UREG var två procent för män och en procent för kvinnor.

Andelen personer som är rätt klassificerade i UREG är lägst för förgymnasial utbildning och ökar sedan med utbildningsnivån. Detta gäller både män och kvinnor. Både antalet kvinnor respektive män med förgymnasial utbildning i UREG är överskattade medan antalet män respektive kvinnor inom övriga utbildningsnivåer är underskattade.

Efter utländsk bakgrund och utbildningsnivå

Utländsk bakgrund Utbildningsnivå	Antal personer enligt;		Andel av UREG;	
	UREG	"sann" kod	Rätt klass	Nettofel
Född i Sverige	5 513 800	5 513 800	87 ± 2	0
Förgymnasial utbildning	1 407 000	1 132 400	78 ± 4	20 ± 4
Gymnasial utbildning	2 547 700	2 655 900	88 ± 2	-4 ± 3
Eftergymnasial utbildning	1 526 200	1 725 500	95 ± 2	-13 ± 3
Okänd utbildningsnivå	32 800	0	0	100
Född utomlands	910 200	910 200	70 ± 5	0
Förgymnasial utbildning	256 300	215 600	68 ± 11	16 ± 13
Gymnasial utbildning	353 800	374 400	72 ± 9	-6 ± 11
Eftergymnasial utbildning	234 200	320 200	89 ± 7	-37 ± 15
Okänd utbildningsnivå	65 800	0	0	100

Andelen personer som har okänd utbildningsnivå i UREG var sju procent för personer födda utomlands och mindre än en procent för personer födda i Sverige. Personer som aldrig har gått i skola redovisas i gruppen okänd utbildningsnivå i UREG. Det finns ingen möjlighet att i UREG se vilka som saknar uppgift om utbildning och vilka som helt saknar utbildning. I evalveringsstudien har det heller inte varit möjligt att ta hänsyn till detta, då de som helt saknar utbildning förmodligen tillhör bortfallet i evalveringsstudien.

Genomförda studier tyder på att folkbokföringen och därmed RTB respektive UREG innehåller ett betydande antal personer som inte längre bor i landet. Under 2004 har studier gjorts där man försökt att skatta storleken på övertäckningen med hjälp av registerstatistik. En preliminär siffra pekar på en övertäckning på minst 35 000 personer.

Det är rimligt att anta att många av dessa tillhör gruppen med okänd utbildningsnivå i UREG och även bortfallet i evalveringsstudien. Detta innebär förmodligen att bruttofelen i UREG blir något överskattade för personer födda utomlands.

Problemen gällande klassificering av utbildningsnivå i UREG verkar vara större för personer födda utomlands än för personer födda i Sverige. Andelen rätt klassificerade är 87 procent för personer födda i Sverige jämfört med 70 procent för personer födda utomlands. Mönstret är dock likartat med överskattning av antalet personer med förgymnasial utbildning och underskattning av antalet personer inom övriga utbildningsnivåer. Störst underskattning tycks finnas av antalet personer födda utomlands med eftergymnasial utbildning.

Efter sysselsättningsstatus och utbildningsnivå

Sysselsättningsstatus Utbildningsnivå	Antal personer enligt;		Andel av UREG;	
	UREG	"sann" kod	Rätt klass	Nettofel
Förvärvsarbetande	4 357 600	4 357 600	85 ± 2	0
Förgymnasial utbildning	814 500	570 500	69 ± 6	30 ± 6
Gymnasial utbildning	2 122 200	2 168 400	85 ± 3	-2 ± 3
Eftergymnasial utbildning	1 410 800	1 618 700	94 ± 2	-15 ± 4
Okänd utbildningsnivå	10 100	0	0	100
Ej förvärvsarbetande	2 066 400	2 066 400	84 ± 3	0
Förgymnasial utbildning	848 900	777 600	84 ± 4	8 ± 5
Gymnasial utbildning	779 300	861 800	89 ± 4	-11 ± 6
Eftergymnasial utbildning	349 600	427 000	95 ± 4	-22 ± 9
Okänd utbildningsnivå	88 500	0	0	100

Andelen personer som har okänd utbildningsnivå i UREG var fyra procent för personer som inte förvärvsarbetade under mätveckan hösten 2002 jämfört med mindre än en procent för personer som förvärvsarbetade.

För både förvärvsarbetande och ej förvärvsarbetande är antalet personer med förgymnasial utbildning överskattade i UREG medan antalet personer inom gymnasial- och eftergymnasial utbildning är underskattade. Överskattningen av antalet personer inom förgymnasial utbildning verkar dock vara större bland de förvärvsarbetande än bland de ej förvärvsarbetande. Andelen rätt klassificerade med förgymnasial utbildning i UREG är också märkbart lägre för förvärvsarbetande jämfört med ej förvärvsarbetande.

Efter frågor i ULF som ligger till grund för kodning av utbildningsnivå. Efter deltagande i kommunal vuxenutbildning och utbildningsnivå

Deltagande i Komvux Utbildningsnivå	Antal personer enligt;		Andel av UREG;	
	UREG	"sann" kod	Rätt klass	Nettofel
Deltagit i Komvux	1 558 000	1 558 000	80 ± 4	0
Förgymnasial utbildning	223 000	114 900	52 ± 13	48 ± 13
Gymnasial utbildning	851 100	863 600	81 ± 5	-1 ± 6
Eftergymnasial utbildning	477 000	579 500	94 ± 4	-21 ± 8
Okänd utbildningsnivå	7 000	0	0	100
Ej deltagit i Komvux	4 866 000	4 866 000	86 ± 2	0
Förgymnasial utbildning	1 440 300	1 233 100	81 ± 4	14 ± 3
Gymnasial utbildning	2 050 500	2 166 700	89 ± 3	-6 ± 4
Eftergymnasial utbildning	1 283 500	1 466 200	94 ± 2	-14 ± 3
Okänd utbildningsnivå	91 700	0	0	100

Andelen personer som har okänd utbildningsnivå i UREG är en procent för personer som deltagit i kommunal vuxenutbildning och två procent för dem som inte deltagit.

Det verkar finnas större problem med klassificeringen av utbildningsnivå bland dem som har deltagit i Komvux än bland dem som inte har deltagit. Detta är speciellt märkbart för personer med förgymnasial utbildning i UREG.

Efter deltagande i universitet/högskola och utbildningsnivå

Deltagande i Universitet/högsk. Utbildningsnivå	Antal personer enligt;		Andel av UREG;	
	UREG	"sann" kod	Rätt klass	Nettofel
Deltagit i Universitet/högskola, Med examen	1 140 700	1 140 700	88 ± 4	0
Förgymnasial utbildning
Gymnasial utbildning
Eftergymnasial utbildning	1 061 000	1 137 700	94 ± 3	-7 ± 3
Okänd utbildningsnivå	20 000	0	0	100
Deltagit i Universitet/högskola, Utan examen	733 500	733 500	72 ± 6	0
Förgymnasial utbildning
Gymnasial utbildning	272 500	143 300	47 ± 10	47 ± 11
Eftergymnasial utbildning	426 700	584 200	93 ± 4	-37 ± 11
Okänd utbildningsnivå	4 600	0	0	100
Ej deltagit i Universitet/högsk.	4 549 800	4 549 800	86 ± 2	0
Förgymnasial utbildning	1 624 500	1 342 100	78 ± 4	17 ± 4
Gymnasial utbildning	2 578 500	2 883 900	92 ± 2	-12 ± 3
Eftergymnasial utbildning	272 700	323 800	97 ± 4	-19 ± 11
Okänd utbildningsnivå	74 100	0	0	100

Andelen personer med okänd utbildningsnivå i UREG är ungefär två procent för personer med examen från universitet/högskola respektive för personer som inte deltagit i utbildning vid universitet/högskola. För personer som inte har examen är andelen ungefär en procent. De största problemen med klassificeringen av utbildningsnivå verkar finnas bland personer som har deltagit i utbildning vid universitet/högskola och inte tagit ut examen. De huvudsakliga källor som kan tänkas fånga upp denna typ av utbildning i UREG är summerade högskolepoäng fr.o.m. 1993 från högskoleregistret samt FoB 1990 och FoB 1970. De som har högskolepoäng från studier tidigare än 1993 fångas således inte upp om de inte har angivit detta i FoB 1990 eller FoB 1970.

Efter deltagande i övrig utbildning och utbildningsnivå. Med övrig utbildning avses utbildningar utanför det reguljära utbildningsväsendet som omfattar minst två månaders studier på heltid

Deltagande i övrig utbildning Utbildningsnivå	Antal personer enligt;		Andel av UREG;	
	UREG	"sann" kod	Rätt klass	Nettofel
Deltagit i övrig utbildning	2 666 900	2 666 900	78 ± 3	0
Förgymnasial utbildning	445 200	166 700	36 ± 8	63 ± 8
Gymnasial utbildning	1 521 900	1 660 400	84 ± 4	-9 ± 4
Eftergymnasial utbildning	695 700	839 800	94 ± 3	-21 ± 6
Okänd utbildningsnivå	4 100	0	0	100
Ej deltagit i övrig utbildning	3 757 100	3 757 100	89 ± 2	0
Förgymnasial utbildning	1 218 200	1 181 300	92 ± 3	3 ± 3
Gymnasial utbildning	1 379 700	1 369 900	89 ± 3	1 ± 4
Eftergymnasial utbildning	1 064 700	1 205 900	94 ± 3	-13 ± 4
Okänd utbildningsnivå	94 500	0	0	100

Andelen personer med okänd utbildningsnivå är tre procent bland personer som inte deltagit i övrig utbildning respektive en procent bland personer som har deltagit i övrig utbildning. Problemen gällande klassificeringen av utbildningsnivå verkar vara större bland personer som har deltagit i övrig utbildning än bland personer som inte har deltagit i övrig utbildning. Speciellt stora problem finns bland personer med förgymnasial utbildning i UREG som har deltagit i övrig utbildning enligt ULF.

Inom övrig utbildning fångas bl.a. kortare yrkesutbildningar och längre personalutbildningar upp, som inte täcks in av UREG.

Skattningar av fel i UREG med avseende på utbildningsinriktning.

Efter källa

Källa	Antal personer enligt; UREG	Andel av UREG;	
		Rätt klass	Bruttofel
Lågt prioriterade källor UREG	471 000	64 ± 6	342 000 ± 49 500
Övriga administrativa register	765 700	69 ± 5	474 000 ± 76 100
FoB 1990	2 297 300	84 ± 3	752 800 ± 112 600
Reguljära statistikregister	2 890 000	94 ± 2	374 300 ± 81 300

Andelen med okänd utbildningsinriktning i UREG är 25 procent för lågt prioriterade källor och nio procent för övriga administrativa register. För övriga källor är andelen mindre än en procent. Andelarna med okänd utbildningsinriktning påverkar storleken på bruttofelen för de olika källorna.

Precis som för utbildningsnivå verkar klassificeringen av utbildningsinriktning i UREG ha fungerat bättre när högsta utbildning är hämtad från reguljära statistikregister än när den är hämtad från FoB 1990.

Efter ålder

Ålder	Antal personer enligt UREG	Andel av UREG;	
		Rätt klass	Bruttofel
16 – 24	937 200	86 ± 4	257 700 ± 61 000
25 – 44	2 431 600	84 ± 3	772 800 ± 102 600
45 – 64	2 314 200	84 ± 3	741 900 ± 114 400
65 – 74	741 000	88 ± 4	170 600 ± 50 600

Andelen med okänd utbildningsinriktning i UREG var åtta procent för personer i åldern 16 – 24 år, fyra procent för åldersgruppen 25 – 44 år, två procent för åldersgruppen 45 – 64 år och en procent för den äldsta åldersgruppen. Bruttofelen tycks relativt sett vara mindre i den äldsta respektive den yngsta åldersgruppen jämfört med åldersgrupperna i mitten. Andelen rätt klassificerade tycks vara högst i den äldsta åldersgruppen följt av den yngsta. Skillnaderna är dock inte signifikanta.

Efter Kön

Kön	Antal personer enligt UREG	Andel av UREG;	
		Rätt klass	Bruttofel
Män	3 235 500	86 ± 2	927 400 ± 118 100
Kvinnor	3 188 500	84 ± 2	1 015 600 ± 125 000

Andelen med okänd utbildningsinriktning var fyra procent för män och tre procent för kvinnor. Det finns inga stora skillnader gällande problem med klassificeringen av utbildningsinriktning mellan kvinnor och män.

Både antalet kvinnor och män med allmän utbildning i UREG är överskattade medan antalet kvinnor och män inom nästan samtliga övriga utbildningsinriktningar är underskattade.

Efter Utländsk bakgrund

Utländsk bakgrund	Antal personer enligt UREG	Andel av UREG;	
		Rätt klass	Bruttofel
Födda i Sverige	5 513 800	87 ± 2	1 405 700 ± 146 100
Födda utomlands	910 200	70 ± 5	537 300 ± 84 100

Andelen personer med okänd utbildningsinriktning i UREG är två procent för personer födda i Sverige och tolv procent för personer födda utomlands. Precis som för utbildningsnivå tycks problemen med klassificeringen av utbildningsinriktning i UREG vara betydligt större för personer födda utomlands än för personer födda i Sverige.

Efter sysselsättningsstatus

Sysselsättningsstatus	Antal personer enligt UREG	Andel av UREG;	
		Rätt klass	Bruttofel
Förvärvsarbetande	4 367 600	85 ± 2	1 287 800 ± 146 300
Ej förvärvs-arbetande	2 056 400	84 ± 3	655 200 ± 96 200

Andelen personer med okänd utbildningsinriktning i UREG är två procent för personer som förvärvsarbetade jämfört med sex procent för personer som inte förvärvsarbetade.

För både personer som är förvärvsarbetande och personer som inte är förvärvsarbetande är antalen med allmän utbildning i UREG överskattade medan antalen inom samtliga övriga utbildningsinriktningar är underskattade. Överskattningen av antalet inom allmän utbildning är större för de förvärvsarbetande än för de ej förvärvsarbetande.

6.2 Orsaker till skillnader mellan Utbildningsregistret och ULF

Under granskningen sattes en kod som angav orsaken till att utbildningsnivå respektive utbildningsinriktning inte matchade mellan UREG och ULF. Koden sattes endast om det gick att bestämma en "sann" kod för utbildningsnivå/utbildningsinriktning.

För utbildningsnivå var det 28 procent av de svarande med "sann" kod som fick en orsakskod. Motsvarande siffra för utbildningsinriktning var 22 procent.

Kartläggningen av orsaker till skillnader mellan UREG och ULF kan hjälpa till vid bedömningen av hur förbättringar kan ske av tillförlitligheten i statistiken från UREG.

6.2.1 Utbildningsnivå

Av de svarande personerna där utbildningsnivå inte matchade mellan UREG och ULF innan granskning saknar 16 procent "sann" kod och 24 procent är kodade enligt UREG. Detta innebär att 60 procent inte är kodade enligt UREG, varav majoriteten är kodade enligt ULF.

Orsaker till skillnader i utbildningsnivå mellan UREG och ULF har blivit skattade med hjälp av orsakskoden.

**Skattning av fördelningen av personer med skillnader i utbildningsnivå mellan UREG och ULF över orsaker till skillnader och källa, procent.
"Sann" kod för utbildningsnivå är satt enligt UREG**

Källa till UREG	Andel personer, % av ej match före granskning			
	Motsägande uppgifter	Utbildning saknas	Ofullständig information	Olika kodningsprinciper
Lågt prioriterade källor UREG	0 ± 0	85 ± 16	7 ± 12	9 ± 12
Övriga administrativa register	2 ± 5	37 ± 14	53 ± 15	8 ± 8
FoB 1990	0 ± 0	52 ± 12	7 ± 7	42 ± 12
Reguljära statistikregister	6 ± 12	0 ± 0	24 ± 21	70 ± 23
Totalt	1 ± 2	48 ± 8	23 ± 7	28 ± 8

De vanligaste anledningarna till att personer har olika utbildningsnivå i UREG och ULF, när den "sanna" koden är satt enligt UREG, är att utbildning saknas i ULF eller att olika principer för kodning har använts i ULF och UREG. Vid olika principer för kodning baseras koden för högsta utbildning på samma utbildning i ULF som i UREG. De principer som har använts i UREG har bedömts vara riktiga.

När högsta utbildning är hämtad från lågt prioriterade källor i UREG är den vanligaste orsaken till skillnader, att utbildning saknas i ULF. Detta beror till stor del på att personer inte uppgett i ULF att de gått på fack- och yrkesskolor på 40- och 50-talet, vilket de har uppgett i FoB 70.

För övriga administrativa register beror brister i ULF till stor del på att informationen från uppgiftslämnarna inte är fullständig eller att utbildningen saknas. Ett exempel på brist i ULF är att kurser som visar sig vara komvux på gymnasienivå enligt komvux-databasen har angetts som "övrig utbildning" i ULF, och därför inte valts som högsta utbildning. När komvux saknas helt i ULF men finns i UREG har posten inte fått någon "sann" kod.

En annan brist i ULF är att personen ofta inte har angett arbetsmarknadsutbildningar, vilka har rapporterats till UREG. I ULF finns ingen specifik fråga om arbetsmarknadsutbildningar utan de skall anges under övriga utbildningar. Det har därför bedömts att utbildningen saknas i ULF.

För FoB 1990 är de största orsakerna till skillnader, att utbildning saknas i ULF eller att andra kodningsprinciper har använts i ULF än i UREG. Ett exempel på när utbildning saknas i ULF är om endast grundskola är angiven i ULF men i FoB 90/UREG har man angett någon kortare gymnasial utbildning. Ett exempel på skillnader i kodningsprinciper är att i ULF sätts gymnasial utbildning 1,5 år till nivå 32, vilket inte överensstämmer med principerna i SUN 2000.

För reguljära statistikregister är den absolut största orsaken till skillnader att andra kodningsprinciper har använts i ULF än i UREG. Detta handlar ofta om att man i ULF låtit avklarade poäng efter examen bli nivåhöjande och att man kodat nivå efter personens alla poäng i en examen medan högskoleregistret kodar nivå efter examenskraven.

I några fall handlar det också om att det behövs mer kunskaper om utbildningarna för att hitta rätt nivå.

Om det inte gått att fastställa nivå utifrån examenskraven för en viss utbildning har det inte gått att sätta en "sann" kod för utbildningsnivå.

"Sann" kod för utbildningsnivå är inte satt enligt UREG

Källa till UREG	Andel personer, % av ej match före granskning			
	Motsägande uppgifter	Utbildning saknas	Ofullständig information	Olika kodningsprinciper
Lågt prioriterade källor UREG	2 ± 4	94 ± 6	3 ± 5	0 ± 0
Övriga administrativa register	3 ± 6	47 ± 17	35 ± 16	16 ± 15
FoB 1990	1 ± 2	81 ± 6	15 ± 6	4 ± 3
Reguljära statistikregister	0 ± 0	62 ± 10	35 ± 10	3 ± 4
Totalt	1 ± 2	75 ± 5	20 ± 4	4 ± 3

Den vanligaste anledningen till att personer har olika utbildningsnivå i UREG och ULF, när den "sanna" koden i de flesta fall är satt enligt ULF, är att utbildning saknas i UREG.

Utländsk utbildning oavsett nivå på utbildningen saknas ibland i UREG.

När högsta utbildning har hämtats från lågt prioriterade källor till UREG är den absolut vanligaste orsaken att den utbildning som anges i ULF saknas i UREG.

För övriga administrativa register är den vanligaste orsaken till skillnader att utbildning saknas i UREG. En annan vanlig orsak är att informationen från källorna till UREG är ofullständig. Ett exempel är att längden på utbildningen saknas.

Den vanligaste orsaken till att posterna ej matchar för källan FoB 90 är att utbildning som anges i ULF saknas i UREG. Det handlar ofta om utbildningar utanför det reguljära utbildningssystemet, t.ex. längre lärlingsutbildningar eller andra icke nivåspecifiserade utbildningar som exempelvis "Bageri 1 år" och "Snickarlärling 4 år". Sådana utbildningar rapporteras inte till UREG på administrativ väg.

En brist i UREG är också att det är vanligt att ange 2-årig gymnasieutbildning (före 1990) i ULF men ej i FoB 90. Utbildningen saknas därför i UREG. Det rör sig troligen om ej avslutade utbildningar eftersom de inte finns i Gymnasieregistren.

För reguljära statistikregister är de vanligaste orsakerna till skillnader i utbildningsnivå mellan UREG och ULF att utbildning saknas i UREG eller att informationen till UREG är ofullständig. Nedan följer några exempel på brister i UREG.

För poster med källan EGGG (ackumulerat gymnasiregister) är det vanligt att man har nivå 32 i UREG men nivå 33 i ULF. I ULF har personer med 2-årig gymnasieutbildning ibland någon ytterligare utbildning/kurs som saknas i UREG, t.ex. "Svets 9 månader". Detta kodas i ULF till nivå 33.

För källan "Högskoleregistret, summerade högskolepoäng" saknas ofta poäng i UREG (summerade högskolepoäng finns endast från 1993 och framåt).

För källan "examina från högskola/universitet" har skillnader i kodningsprinciper satts som orsak när UREG använt uppgifter som översatts från gamla SUN till SUN2000. I flera fall finns "ULF:s" nivå i högskoleregistret. Utbildning saknas i UREG när examina från högskola/universitet är avlagda under andra halvåret 2002. Uppgifter om påbyggnadsutbildningar för sjuksköterskor saknas i högskoleregistret, eftersom det under större delen av 1990-talet inte utfärdades några examina för dessa utbildningar.

6.2.2 Utbildningsinriktning

Av de svarande personerna där utbildningsinriktning inte matchade mellan UREG och ULF innan granskning saknar 14 procent "sann" kod och 27 procent är kodade enligt UREG. Detta innebär att 59 procent inte är kodade enligt UREG, varav majoriteten är kodade enligt ULF.

Orsaker till skillnader i utbildningsinriktning mellan UREG och ULF har blivit skattade med hjälp av orsakskoden.

Skattning av fördelningen av personer med skillnader i utbildningsinriktning mellan UREG och ULF över orsaker till skillnader och källa, procent. "Sann" kod för utbildningsinriktning är satt enligt UREG

Källa till UREG	Andel personer, % av ej match före granskning			
	Motsägande uppgifter	Utbildning saknas	Ofullständig information	Olika kodningsprinciper
Lågt prioriterade källor UREG	2 ± 5	81 ± 15	6 ± 11	10 ± 12
Övriga administrativa register	7 ± 12	50 ± 21	23 ± 18	21 ± 17
FoB 1990	0 ± 0	43 ± 13	8 ± 7	50 ± 13
Reguljära statistikregister	0 ± 0	7 ± 7	5 ± 6	88 ± 9
Totalt	1 ± 2	39 ± 8	8 ± 5	51 ± 8

De vanligaste anledningarna till skillnader mellan ULF och UREG, när "sann" kod är satt enligt UREG, är olika kodningsprinciper eller att utbildning saknas i ULF. När utbildning saknas blir både utbildningsnivå och utbildningsinriktning i de flesta fall felaktiga i ULF. De orsaker som är redovisade under skillnader i utbildningsnivå leder därför i de flesta fall också till skillnader i utbildningsinriktning.

Den största skillnaden mellan fördelningen över orsakskoder vid jämförelse mellan utbildningsnivå och utbildningsinriktning är att det är mer vanligt med olika kodningsprinciper för utbildningsinriktning än för utbildningsnivå. Detta beror på att samma utbildning har blivit redovisad i ULF som i UREG men de har fått olika koder för utbildningsinriktning. Ett exempel är att samhällsvetenskapligt och naturvetenskapligt program på gymnasiet i ULF ofta kodas med andra inriktningar än '0', vilket är en kodningsbrist i ULF.

”Sann” kod för utbildningsinriktning är inte satt enligt UREG

Källa till UREG	Andel personer, % av ej match före granskning			
	Motsägande uppgifter	Utbildning saknas	Ofullständig information	Olika kodningsprinciper
Lågt prioriterade källor UREG	2 ± 5	91 ± 8	0 ± 0	6 ± 7
Övriga administrativa register	3 ± 4	46 ± 12	30 ± 10	21 ± 9
FoB 1990	0 ± 0	95 ± 4	2 ± 3	3 ± 3
Reguljära statistikregister	4 ± 5	76 ± 10	8 ± 7	11 ± 8
Totalt	2 ± 2	79 ± 4	9 ± 3	9 ± 4

Den vanligaste anledningen till skillnader i utbildningsinriktning mellan ULF och UREG, när den ”sanna” koden i de flesta fall är satt enligt ULF, är att utbildning saknas i UREG. När utbildning saknas blir både utbildningsnivå och utbildningsinriktning i de flesta fall felaktiga i UREG. De orsaker som är redovisade under skillnader i utbildningsnivå leder därför i de flesta fall också till skillnader i utbildningsinriktning.

Ofullständig information är mindre vanligt som orsak till skillnader i utbildningsinriktning än som orsak till skillnader i utbildningsnivå. Detta beror förmodligen på att den ofullständiga informationen i UREG i högre grad har gällt längden på utbildningen än huvudsakligt innehåll i utbildningen.

Ett exempel på olika kodningsprinciper är att schablonkodning av komvux-inriktning i UREG inte alltid blir korrekt.

Bilaga 1

Skattningsförfarande i evalveringsstudien

Osäkerhetskällor

ULF är en urvalsundersökning och detta innebär att skattningarna i evalveringsstudien är behäftade med urvalsfel beroende på att endast en delmängd (urval) av populationen studeras. Ett annat fel uppkommer om personer som det saknas svar för (bortfall) avviker från de svarande med avseende på undersökningsvariablerna, exempelvis om utbildningsnivå är rätt klassificerad eller inte. Detta fel kallas för bortfallsfel. I evalveringsstudien består bortfallet dels av personer som saknar svar från ULF och dels av personer som inte har tilldelats någon "sann" kod för utbildningsnivå/utbildningsinriktning.

Registret över totalbefolkningen (RTB) utgör urvalsram för ULF. I RTB förekommer i viss mån täckningsfel, vilka består i att personer som är stadigvarande bosatta i Sverige inte är folkbokförda (undertäckning) samt att personer är folkbokförda trots att de inte är bosatta i Sverige (övertäckning). Vid framtagningarna av skattningarna av fel i UREG har de totaler som används för att fastställa de kalibrerade vikterna hämtats från UREG version 2003-01-01. Detta förfarande bygger på ett modellantagande om hur undersökningsvariablerna är relaterade till kalibreringsvariablerna.

I evalveringsstudien finns också mätfel, då de "sanna" värdena för utbildningsnivå respektive utbildningsinriktning endast kan betraktas som approximativa värden. Mätfel kan ge systematiska fel i skattningarna och orsaker till detta kan vara felaktigt lämnade svar i ULF eller UREG samt att felaktiga kriterier har använts vid bestämningen av "sanna" värden.

Kalibrering

I evalveringsstudien antas att urvalet har dragits med stratifierat obundet slumpmässigt urval (STOSU) vid urvalstillfället. Kön, åldersklasser, H-region, civilstånd och fått lön under året bildar 96 stratum. Vid framtagningen av skattningarna till ULF delas i själva verket population och urval in i 96 grupper (poststrata) vid estimationstillfället istället för vid urvalstillfället. Det finns hjälpinformation från UREG som kan antas vara effektiv vid estimationen. Detta estimationsförfarande, som utnyttjar hjälpinformation i form av registervariabler, används i det här fallet främst för att kompensera för den skevhet som kan antas uppstå till följd av bortfall. Förfarandet kan också reducera den slumpmässiga osäkerheten (variansen) som skattningarna är behäftade med.

Kalibreringstekniken innebär att registervariabler (hjälpvariabler) används för att "vikta upp" underrepresenterade grupper. Om t.ex. personer med kort utbildning är underrepresenterade bland de svarande och utbildningsnivå används i kalibreringen kommer personer med kort utbildning att få större vikt än personer med lång utbildning.

De kalibrerade vikterna har den egenskapen att de resulterar i en skattning som överensstämmer exakt med den registertotal som används då de kalibrerade vikterna beräknas. Detta innebär bl.a. att skattningar av antal personer för kalibreringsvariablerna kommer att överensstämma med de storheter som kan beräknas utifrån UREG version 2003-01-01.⁴

Det centrala arbetet för att få god kvalitet på skattningarna, då kalibrerings-estimatoren används, är att använda "stark" hjälpinformation. I följande avsnitt beskrivs detta arbete för evalveringsstudien.

Hjälpvariabler i evalveringsstudien

Var och en av de hjälpvariabler som används i kalibreringen har valts utifrån minst ett av följande kriterier (se Lundström S och Särndal C (2001)):

- i) De samvarierar med svarsbenägenheten.
- ii) Samvarierar med (viktiga) undersökningsvariabler.
- iii) Avgränsar (viktiga) redovisningsgrupper

Tänkbara hjälpvariabler, det vill säga variabler som tros uppfylla åtminstone ett av ovan uppsatta kriterier, hämtades ifrån UREG version 2003-01-01 och RTB. I det här fallet har främst punkterna (i) och (iii) varit vägledande. Om hjälpvariablerna samvarierar med svarsbenägenheten kan man förvänta sig att användandet av kalibrerade vikter leder till en reduktion av bortfallsskevheten. Är punkt (ii) uppfylld bör såväl variansen (samplingvarians + bortfallsvarians) som bortfallsskevheten reduceras för skattningar baserade på de viktiga undersökningsvariablerna.

Felen i utbildningsinriktning respektive utbildningsnivå skattas var för sig med olika uppsättningar av hjälpvariabler, vilket innebär att två olika kalibreringsestimatörer används.

För utbildningsnivå har följande hjälpvariabler använts;

- Utbildningsinriktning enligt huvudinriktning SUN 2000 (10 grupper)
- Utbildningsnivå på aggregerad tvåsiffer-nivå enligt SUN 2000 (7 grupper)
- Källa till UREG (lågt prioriterade källor till UREG, övriga administrativa register, reguljära statistikregister och FoB 1990) kombinerat med utbildningsnivå (förgymnasial, gymnasial, eftergymnasial). Kombination med utbildningsnivå görs endast för reguljära statistikregister och FoB 90 (8 grupper)
- Ålder (16-24, 25-44, 45-64, 65-74) kombinerat med utbildningsnivå (förgymnasial, gymnasial, eftergymnasial). Detta resulterar i tolv grupper.
- Kön kombinerat med utbildningsnivå (6 grupper)
- Födelseland (Utrikes födda respektive födda i Sverige) kombinerat med utbildningsnivå (6 grupper)

⁴ Teknisk rapportmall 2003-12-21, BV/ENK

För utbildningsinriktning har följande hjälpvariabler använts;

- Utbildningsinriktning enligt huvudinriktning SUN 2000 (10 grupper)
- Utbildningsnivå på aggregerad tvåsiffer-nivå enligt SUN 2000 (7 grupper)
- Källa till UREG kombinerat med utbildningsinriktning. Utbildningsinriktning bildas av Allmän utbildning+tjänster+okänd utbildning, Humaniora och konst+samällsvetenskap, Handel och administration, Naturvetenskap+Teknik och Tillverkning+Lant-och skogsbruk samt Pedagogik och lärarutbildning+Hälsa- och sjukvård, vilket blir fyra grupper. Detta resulterar i 16 grupper.
- Ålder (16-24, 25-44, 45-64, 65-74) kombinerat med utbildningsinriktning. Detta resulterar i 16 grupper.
- Kön kombinerat med utbildningsinriktning (8 grupper)
- Födelseland (Utrikes födda respektive födda i Sverige) kombinerat med utbildningsinriktning (8 grupper)

I en studie av bortfallet för 2000 års ULF⁵ visade det sig som tidigare nämnts att bortfallet i ULF snedvred några skattningar av registervariabler. De variabler som utifrån detta tycks viktiga att ta med är födda i Sverige, födda utomlands, utbildningsnivå och sysselsatt eller inte. I evalveringsstudien är bortfallet lite högre än i ULF eftersom det inte har varit möjligt att sätta "sann" kod för samtliga svarande i ULF. En studie av bortfallets fördelning efter utrikes födda/födda i Sverige, utbildningsnivå och sysselsatt/inte sysselsatt visade dock att mönstret såg likadant ut i evalveringsstudien som i ULF. Felen i UREG kan också antas vara större för utrikes födda än för inrikes födda eftersom andelen med okänd utbildning är betydligt större bland utrikes födda. Vilken källa till UREG som har använts för att hämta högsta utbildning kan också antas ha en effekt på storleken på felet.

Utbildningsnivå och utbildningsinriktning är de absolut viktigaste redovisningsgrupperna för att skatta felet i uppgifterna från UREG. Andra redovisningsgrupper som bedömdes viktiga var källor till UREG, ålder, kön och född i Sverige/ född utomlands.

Ovanstående resonemang ledde fram till de uppsättningar av hjälpvariabler som användes för utbildningsnivå respektive utbildningsinriktning.

Redovisning av osäkerhetsmått

Osäkerhetsintervall beräknas enligt formeln för nominellt 95-procentiga konfidensintervall:

punktskattning $\pm 1,96$ *(skattning av estimatorns medelfel)

Under förutsättning att stickprovsstorleken är stor samt att

- 1) estimatorn är approximativt väntevärdesriktig, dvs. icke-skev
- 2) estimatorns varians beror endast på sampling- och bortfallsvariansen
- 3) hjälpinformationen är starkt korrelerad med de okända svarssannolikheterna

⁵ Bakgrundsfakta 2003:8, Hassan Mirza

gäller att intervallet har en faktisk konfidensgrad på ca 95 procent. Detta innebär att om dessa antaganden är uppfyllda är sannolikheten ca 95 procent att det använda urvals- och skattningsförfarandet leder till ett intervall som täcker in den sökta storheten.

Dessa förutsättningar är inte perfekt uppfyllda i evalveringsstudien. Även om hjälpvariablerna är klokt valda kommer estimatorn att vara behäftad med viss skevhet, delvis till följd av kvarstående bortfalls-, täcknings- och mätfel av systematisk karaktär. Vidare är det troligt att estimatorns varians underskattas. Utöver sampling- och bortfallsvariansen tillkommer även variansbidrag från andra osäkerhetskällor av slumpmässig natur. Det är inte möjligt att avgöra vilka effekter avvikelserna från förutsättningarna ovan får för tolkningen av de osäkerhetsintervall som presenteras i resultaten. En rekommendation är att de presenterade intervallen inte tolkas som regelrätta konfidensintervall.

Teknisk beskrivning av urval och estimation

Punktskattningar

Från en population U bestående av N individer dras ett slumpmässigt urval s av storleken n_s enligt designen $p(\cdot)$ som är sådan att samtliga individer i populationen har en sannolikhet > 0 att komma med i urvalet. På grund av bortfall kan data om y -variabeln samlas in endast för en delmängd r av storleken m_r .

I evalveringsstudien innebär designen $P(\cdot)$ att populationen antas delas in i H strata innan urvalsdragning, där strata h innehåller N_h individer. Det antas vidare att ett slumpmässigt urval av storleken n_h har dragits inom varje stratum h så att alla individer hade samma sannolikhet att komma med i urvalet. De strata som används i evalveringsstudien sammanfaller i själva verket med de 96 grupper (poststrata) som används vid estimations-tillfället för framtagningen av skattningarna från ULF. I evalveringsstudien har bortfall uppstått, dels p.g.a. bortfall i ULF och dels p.g.a. att det inte har gått att bestämma "sann" kod för utbildningsnivå/utbildningsinriktning vid granskningen. Detta innebär att det endast finns svar från mängden r_h inom varje stratum.

De parametrar vi är intresserade av är främst funktioner av två totaler $Y = \sum_U y_k$ och $Z = \sum_U z_k$, där y_k är värdet på variabel y för person k och z_k värdet på en annan variabel för samma person. I evalveringsstudien definierar vi y och z som dikotoma variabler, enligt följande;

$$y_k = \begin{cases} 1 & \text{om för person } k; \text{ nivå/inriktning i UREG} = \text{"sann"} \text{ nivå/inriktning} \\ 0 & \text{annars} \end{cases} \quad (5.1)$$

$$z_k = 1$$

Det finns givetvis också intresse av parametrar för olika redovisningsgrupper framförallt efter utbildningsnivå/utbildningsinriktning. Låt oss exempelvis benämna utbildningsnivå/utbildningsinriktning enligt UREG = $U_1, \dots, U_j, \dots, U_J$ respektive "sann"

utbildningsnivå/utbildningsinriktning enligt evalveringsstudien =

$$U_1, \dots, U_i, \dots, U_I \text{ där } U = \bigcup_{j=1}^J U_j \text{ och } U = \bigcup_{i=1}^I U_i.$$

Totala antalet rätt klassificerade för en viss utbildningsnivå/utbildningsinriktning enligt UREG, j , och evalveringsstudien, i , kan skrivas

$$Y_{ij} = \sum_U y_{ijk} = t.ex. N_{11} \quad (5.2)$$

$$\text{där } y_{ijk} = \begin{cases} y_k \text{ för } k \in U_j \text{ och } U_i \\ 0 \text{ annars} \end{cases}$$

Totala antalet personer inom en viss utbildningsnivå/utbildningsinriktning enligt UREG, j , kan skrivas;

$$Z_j = \sum_U z_{jk} = t.ex. N_{.1}$$

$$\text{där } z_{jk} = \begin{cases} z_k \text{ för } k \in U_j \\ 0 \text{ annars} \end{cases}$$

Totala antalet personer inom en viss "sann" utbildningsnivå/utbildningsinriktning enligt evalveringsstudien, i , kan skrivas;

$$Z_i = \sum_U z_{ik} = t.ex. N_{1.}$$

$$\text{där } z_{ik} = \begin{cases} z_k \text{ för } k \in U_i \\ 0 \text{ annars} \end{cases}$$

En generell parameter för redovisningsgrupperna i och j (i och j kan också avse hela populationen) kan skrivas $\theta_{ij} = C \frac{Y_{ij}}{Z_j}$, där C är en konstant.

Parametrar av intresse för evalveringsstudien har blivit beskrivna i avsnitt 5.2.1. Andelen rätt klassificerade erhålles när $C = 100$ enligt följande;

$$ARK_{ij} = 100 \frac{Y_{ij}}{Z_j} = t.ex. 100 \frac{N_{11}}{N_{.1}} \text{ när } j = 1 \text{ och } i = 1 \quad (5.3)$$

Relativa nettofel respektive totala antalet bruttofel definieras enligt följande;

Andel relativa nettofel =

$$RNF_{ij} = 100 \frac{Z_j - Z_i}{Z_j} = t.ex. 100 \frac{N_{.1} - N_{1.}}{N_{.1}} \text{ när } j=1 \text{ och } i=1$$

Totala antalet bruttofel = $TBF_{ij} = Z_j + Z_i - 2Y_{ij} = t.ex. N_{.1} + N_{1.} - 2N_{11}$
när $j=1$ och $i=1$

En skattning av totalen Y_{ij} baserad på de svarande personerna i evalveringsstudien definieras som:

$$\hat{Y}_{ij} = \sum_{h=1}^{96} \frac{N_h}{m_h} \sum_{r_h} y_{ijk} = \sum_{h=1}^{96} N_h \bar{y}_{ijr_h} \quad (5.4)$$

Där

m_h = Antalet svarande personer inom stratum h (som i själva verket sammanfaller med antal svarande inom poststrata g som används för framtagningarna av skattningarna från ULF).

I syfte att erhålla en estimator med mindre urvalsfel och bortfallskevheter än estimator (5.4) skattas totalen Y_{ij} med en kalibreringsestimator som har följande utseende:

$$\hat{Y}_{w_{ij}} = \sum_r d_k v_k y_{ijk} \quad (5.5)$$

där

$$d_k = N_h / n_h \text{ för } k \in r_h$$

och

$$v_k = 1 + \left(\sum_U \mathbf{x}_k - \sum_r d_k \mathbf{x}_k \right) \left(\sum_r d_k \mathbf{x}_k \mathbf{x}_k' \right)^{-1} \mathbf{x}_k \quad (5.6)$$

v_k kan uppfattas som en korrektionsfaktor som är hjälpinformationens bidrag för att minska bortfalls- och slumpfel.

$\mathbf{x}_k = (x_{1k}, \dots, x_{jk}, \dots, x_{Jk})$ är en vektor av längden J. Där J är antalet hjälpvariabler.

$\mathbf{t}_x = \sum_U \mathbf{x}_k = (t_{x1}, \dots, t_{xj}, \dots, t_{xJ})$ är en vektor av längden J som innehåller kända totaler från register.

$\hat{\mathbf{t}}_x = (\hat{t}_{x1}, \dots, \hat{t}_{xj}, \dots, \hat{t}_{xJ})$ är en vektor som innehåller skattningar på elementen i \mathbf{t}_x . Där skattningen av varje element t_{xj} ges av: $\hat{t}_x = \sum_r d_k x_k$

Vid skattning av en parameter av typen $\theta_{ij} = C \frac{Y_{ij}}{Z_j}$ skattas respektive total

med hjälp av kalibreringsvikterna $w_k = d_k v_k$ som beror av den antagna urvalsdesignen och hjälpvektorn \mathbf{x}_k . Kalibreringsvikten, v_k , justerar för bortfallet, vilket baseras på antagandet om stark korrelation mellan hjälpvariablerna och de okända svarssannolikheterna.

Kalibreringsvikten uppfyller villkoret: $\sum_r w_k \mathbf{x}_k = \sum_u \mathbf{x}_k$, vilket innebär att om vikterna läggs på variabler som ingår i hjälpvektorn summeras dessa upp till de hjälptotaler vi hämtat från registren.

Variansskattningar

Den varians estimator som används har härletts utifrån den två-fas metod som finns beskriven i Särndal, Swensson och Wretman (1992) tillsammans med antagandet att individerna i stickprovet svarar oberoende av varandra.

Varians estimatorn, här för totalskattningen av Y , kan delas upp i två termer, $\hat{V}(\hat{Y}_{wij}) = \hat{V}_{SAM} + \hat{V}_{NR}$. Komponenten \hat{V}_{SAM} är den första fasens varianskomponent och skattar variansbidraget från urvalsselekteringen medan \hat{V}_{NR} är den andra fasens varianskomponent och skattar variansbidraget från selekteringen till antalet svarande.

Variansen för \hat{Y}_{wij} skattas med

$$\hat{V}_{SAM} = \sum_h \frac{N_h^2}{n_h} \left(1 - \frac{n_h}{N_h}\right) \left[\frac{1}{n_h - 1} \left\{ \sum_{r_h} (g_k v_{sk} e_k)^2 - \frac{(\sum_{r_h} g_k v_{sk} e_k)^2}{n_h} \right\} - \frac{1}{n_h} \sum_{r_h} \left(1 - \frac{1}{v_{sk}}\right) (g_k v_{sk} e_k)^2 \right], \quad (5.7)$$

$$\hat{V}_{NR} = \sum_h \frac{N_h^2}{n_h^2} \sum_{r_h} \left(1 - \frac{1}{v_{sk}}\right) (v_{sk} e_k)^2. \quad (5.8)$$

där e_k är den skattade residualen $e_k = y_k - \mathbf{x}'_k \hat{\mathbf{B}}$, $k \in r$,

$$\hat{\mathbf{B}} = \left(\sum_h \frac{N_h}{n_h} \sum_{r_h} v_{sk} \mathbf{x}_k \mathbf{x}'_k \right)^{-1} \sum_h \frac{N_h}{n_h} \sum_{r_h} v_{sk} y_k \mathbf{x}_k$$

och

$$g_k = 1 + \left(\mathbf{t}_x - \sum_h \frac{N_h}{n_h} \sum_{s_h} \mathbf{x}_k \right)' \left(\sum_h \frac{N_h}{n_h} \sum_{s_h} \mathbf{x}_k \mathbf{x}_k' \right)^{-1} \mathbf{x}_k$$

och

$$v_{sk} = 1 + \left(\sum_h \frac{N_h}{n_h} \sum_{s_h} \mathbf{x}_k - \sum_h \frac{N_h}{n_h} \sum_{r_h} \mathbf{x}_k \right)' \left(\sum_h \frac{N_h}{n_h} \sum_{r_h} \mathbf{x}_k \mathbf{x}_k' \right)^{-1} \mathbf{x}_k$$

De redovisade osäkerhetsmått (95 %-iga konfidensintervall) definieras som

$$1,96 * \sqrt{\hat{V}(\hat{y}_{wij})}$$

Alla beräkningar av punkt- och variansskattningar för andelen rätt klassificerade, relativt nettofel och totalt bruttofel har utförts med programpaketet CLAN97 i SAS.

2004:1 Örestat II, Utvidgad och integrerad flyttningsstatistik, Förstudie
2004:2 Frågebank över bakgrundsfrågor i postenkätundersökningar
2004:3 Samordning av BHU-H och HEK – Samordning 2003
2004:4 Könsuppdelad statistik – Återrapportering av ett uppdrag i regleringsbrevet
2004:5 Flergenerationsregistret 2003 – En beskrivning av innehåll och kvalitet
2004:6 Den nya HUT:en, Hushållens utgifter
2004:7 Multi-Generation Register 2003 - A description of contents and quality
2005:1 Flergenerationsregistret 2004 – En beskrivning av innehåll och kvalitet
2005:2 Multi-Generation Register 2004 - A Description of contents and quality
2005:3 Datafångst av personakter för komplettering av Flergenerationsregister
2005:4 Historiska befolkningsregistret (HBR)
2005:5 Migrations- och integrationsstatistik vid SCB – nuläge och utveckling
2006:1 Publikationsprojektet – slutrapport
2006:2 Översyn av forskarutbildningsstatistiken
2006:3 Historic population register

ISSN 1650-4917
ISBN 91-618-1334-6
ISBN 978-91-618-1334-6