

Nya metoder för att hantera avaktualisering av okänd orsak

© Arbetsförmedlingen
Författare: Gisela Waisman
Datum: 2022-02-25
Diarienummer: Af-2021/0049 6067
Arbetsförmedlingen analys 2022:5

Innehåll

Sammanfattning	4
1 Inledning	5
1.1 En del arbetssökande lämnar Arbetsförmedlingen utan att säga varför	5
1.2 Avaktualiseringsorsaken är ett viktigt utfall i många utvärderingar	5
1.3 Tidigare lösning på problemet	6
1.4 Rapportens syfte	6
1.5 Nytt sätt att lösa problemet	6
1.6 Resultat	7
1.7 Rapportens disposition	7
2 Tidigare studier	8
3 Data	9
4 Metoder för att välja delgrupper	13
5 Hur påverkas sysselsättningsgraden av individers egenskaper?	16
5.1 Utbildningsnivåer	16
5.2 Uppgift om eventuell funktionsnedsättning	17
5.3 Kvinnor och män	17
5.4 Ålder	18
5.5 Inrikes och utrikes födda	18
6 Indelning i delgrupper med ett beslutsträd	19
6.1 Indelning i fyra delgrupper	19
6.2 Indelning i fem delgrupper	20
6.3 Indelning i sex delgrupper	22
6.4 Resultat av beslutsträdsanalysen	22
7 Indelning i delgrupper med ett gränsvärde för skillnad i sysselsättningsgrad	23
7.1 Kombinationer av utbildningsnivå och kod för funktionsnedsättning	23
7.2 Indelning av kombinationer av utbildningsnivå och uppgift om eventuell funktionsnedsättning	24
7.3 Resultat av analysen med hjälp av ett gränsvärde för skillnad i sysselsättningsgrad	25
8 Slutsats	26
Referenser	28
Bilaga 1: Jämförelse av klassificering av utbildning i databasen LISA och i Arbetsförmedlingens Datalagret	29
Bilaga 2: Uppdateringen av systemet för klassificering av utbildning	31
Förändringar i SUN 2020	31
Effekt av uppdateringen i SUN 2020	31
Bilaga 3: Analys av månadsdata år 2019	34

Sammanfattning

Att löpande utvärdera arbetsmarknadspolitiska program och insatser är centralt för att skapa kunskap som bidrar till att rätt insats kan ges till rätt individ i rätt tid. För att kunna genomföra dessa utvärderingar behövs information om i vilken utsträckning arbetssökande övergått till arbete efter att ha deltagit i ett program eller en insats.

I Arbetsförmedlingens datalager dokumenteras orsaken till att arbetssökande inte längre är inskrivna som arbetssökande på Arbetsförmedlingen, så kallad avaktualiseringsorsak. Möjliga orsaker är att personen fått ett arbete eller påbörjat en utbildning. Samtidigt lämnar många arbetssökande Arbetsförmedlingen utan att informera om varför, vilket registreras som avaktualisering av okänd orsak. En del av dessa arbetssökande kan ha fått ett arbete.

Syftet med denna rapport är att redovisa hur man i analyser och utvärderingar kan öka kvaliteten i uppgifterna om orsaken till att arbetssökande inte längre är inskrivna på Arbetsförmedlingen.

Avsaknaden av information om varför en arbetssökande avslutat en arbetslöshetsperiod kan innebära att andelen som har fått jobb, till exempel efter ett arbetsmarknadspolitiskt program, antas vara lägre än vad den är i verkligheten. Underskattningen av andelen som får ett arbete kan snedvrida resultat av utvärderingar av effekten av programmet. Även andra analyser kan snedvridnas om andelen som har fått arbete underskattas.

I rapporten presenteras två olika metoder för att komplettera de ofullständiga registeruppgifterna om varför arbetssökande inte längre är inskrivna på Arbetsförmedlingen. Båda metoder delar in personer med avaktualisering av okänd orsak i delgrupper som skiljer sig tillräckligt mycket för att det ska vara meningsfullt att inkludera information per delgrupp. I den första metoden används ett så kallad beslutsträd medan i den andra metoden används ett gränsvärde för att skapa mindre grupper med lika karakteristika.

Genom att använda någon av de alternativa metoderna är det möjligt att förbättra kvaliteten i såväl enklare uppföljningar, analyser som utvärderingar. Vilket alternativ som väljs kommer att bero på behovet av information och kostnaden för att inkludera mer information i analysen.

1 Inledning

Arbetsförmedlingens uppdrag handlar i grunden om att bidra till en väl fungerande arbetsmarknad. För att veta hur väl Arbetsförmedlingen utför sitt uppdrag behövs kunskap om de resultat och effekter som verksamheten leder till. Ett viktigt resultat är hur många personer som övergår till arbete efter ett arbetsmarknadspolitiskt program eller en insats.

1.1 En del arbetssökande lämnar Arbetsförmedlingen utan att säga varför

I Arbetsförmedlingens register dokumenteras varför arbetssökande avslutat en arbetslöshetsperiod, så kallad avaktualiseringsorsak. Orsaken kan till exempel vara att personen har fått ett arbete eller påbörjat en utbildning. Problemet är att många arbetssökande inte upprätthåller kontakten med Arbetsförmedlingen efter avslutad arbetslöshetsperiod vars avslut kan bero på att de fått ett arbete.

Arbetsförmedlingen försöker alltid få vetskap om varför en arbetssökande inte längre önskar vara inskriven. Detta för att avaktualiseringen ska kunna ske med rätt orsak. En skriftlig förfrågan skickas när Arbetsförmedlingen har försökt få kontakt med den arbetssökande via telefon men inte lyckats. Av texten i förfrågan framgår det när den arbetssökande senast måste kontakta Arbetsförmedlingen för att inte bli avaktualiserad. Om den arbetssökande inte besvarar förfrågan inom utsatt tid blir det en avaktualisering av okänd orsak.

1.2 Avaktualiseringsorsaken är ett viktigt utfall i många utvärderingar

I utvärderingar av arbetsmarknadspolitiska program jämförs ofta andelen som övergår till arbete bland en grupp deltagare i ett program med en grupp av individer som inte deltagit i programmet. Vi kan inte veta om personer som lämnade Arbetsförmedlingen av okänd orsak har övergått till arbete eller inte. Om bortfallet inte hanteras underskattas andelen som har fått jobb i både deltagar- och jämförelsegrupp vilket kan snedvrida utvärderingarnas resultat.

Problemet med bortfall accentueras av att andelen som lämnat Arbetsförmedlingen av okänd orsak brukar skilja sig åt mellan deltagar- och jämförelsegrupp. Bland programdeltagarna har ofta färre lämnat av okänd orsak än i jämförelsegruppen. Det kan bero på att förmedlarna har närmare kontakt med deltagarna än med övriga inskrivna och därmed mer information om anledningen till att deltagarna lämnar Arbetsförmedlingen.

1.3 Tidigare lösning på problemet

I tidigare effektutvärderingar har det antagits att de som lämnat Arbetsförmedlingen av okänd orsak har fått arbete med 50 procent sannolikhet. Det gäller för alla program och för både deltagar- och jämförelsegrupp. Antagandet är rimligt för program vars deltagare och jämförelsepersoner har liknande egenskaper som genomsnittet av de som avaktualiseras av okänd orsak. Det är mindre rimligt för program vars deltagare står närmare eller längre ifrån arbetsmarknaden än genomsnittet.

Tidigare studier har visat att det finns egenskaper, till exempel om en person har en viss utbildningsnivå eller ålder, som påverkar sannolikheten att ha fått arbete bland de som avaktualiserats av okänd orsak. Deltagarna (och jämförelsepersonerna) inom olika program skiljer sig åt vad det gäller dessa egenskaper och därför stämmer antagandet om 50 procent bättre för vissa program än för andra.

Arbetsförmedlingen (2016) visar att en imputeringsmodell som använder registerdata från 2011/2012 har 74 procent korrekta prediktioner, medan slumpmässig imputering inom hela gruppen (motsvarande den metoden som har använts i tidigare utvärderingar) ger 53 procent korrekta prediktioner. Ett möjligt sätt att hantera bortfallet är att använda uppskattningar baserade på resultat från en sådan imputeringsmodell. Problemet med det sättet att hantera bortfallet är att man skulle behöva inkludera en imputeringsmodell i varje effektutvärderingsprogram. En modell som är lika omfattande som den i Arbetsförmedlingen (2016) skulle kräva omfattande bearbetning av data. En modell med färre variabler skulle vara lättare att använda men också minska prediktionsförmågan.

1.4 Rapportens syfte

Syftet med analysen är att förbättra utvärderingarnas träffsäkerhet genom att hitta ett enkelt sätt att hantera bortfallet i utvärderingar och därmed minska snedvridning av resultat. Vi söker en metod för att indela de personer som har lämnat Arbetsförmedlingen utan att säga varför i mindre grupper med liknande sannolikhet att ha fått ett arbete. Metoden måste vara enkel nog för att resultaten ska lätt kunna användas i analyser av andel som har fått ett arbete efter viss tid eller efter viss behandling som till exempel ett arbetsmarknadspolitiskt program.¹

1.5 Nytt sätt att lösa problemet

De personer som lämnat Arbetsförmedlingen av okänd orsak delas in i delgrupper utifrån de egenskaper som har störst betydelse för deras jobbchanser. Med hjälp av dessa delgrupper kommer framtida utvärderingar att kunna anpassa antagandet om vilken andel som fått arbete av de som avregistrerats av okänd orsak till just deras

¹ Resultaten ska också lätt kunna inkluderas i programscript för effektutvärdering utan att belasta scripten med för mycket kod och göra datakörningarna för långsamma.

deltagar- och jämförelsegrupper. Då minskar snedvridning av utvärderingens resultat.

Även fortsättningsvis kommer andelen som avaktualiseras av okänd orsak att skilja sig åt mellan deltagar- och jämförelsegrupp i utvärderingar, men skillnaden och dess påverkan på effekten blir mindre problematiska ju bättre bortfallet hanteras.

Analysen baseras på sysselsättningsstatus av de personer som avaktualiserades av okänd orsak under perioden 2015 – 2019 uppskattade med hjälp av SCB:s Statistiska centralbyråns (SCB) registerdata. Dessa personer kan delas in i delgrupper efter egenskaper såsom utbildningsnivå, uppgift om eventuell funktionsnedsättning, kön, ålder och födelseland. Två indelningsmetoder används och analyseras: först används en statistisk metod och sedan ett valt gränsvärde för hur små skillnader kan vara.

I den första metoden används den statistiska metoden CART (Classification and Regression Tree) för att lösa klassificeringsproblemet med hjälp av ett beslutsträd. Den statistiska metoden indikerar att få delgrupper (mellan fyra och sex delgrupper) ger en bra indelning av de personer som lämnade Arbetsförmedlingen av okänd orsak.

Sedan används ett gränsvärde för att uppnå så kallade ”relevanta” delgrupper. Med relevanta delgrupper menas här att dessa delgrupper skiljer sig tillräckligt mycket i sysselsättningsgrad för att det ska vara meningsfullt att ha ett antagande om sysselsättningsgrad per delgrupp i stället för ett generellt antagande för hela gruppen. Med hjälp av ett valt gränsvärde för hur små skillnader kan vara delas personer som avaktualiserades av okänd orsak i elva relevanta delgrupper.

1.6 Resultat

Analysen resulterar i fyra alternativa indelningar att välja ifrån för att kunna hantera bortfallet när avaktualiseringsorsaken är okänd. I varje effektutvärdering som använder avaktualiseringsorsak som utfall kan man välja en mer eller mindre detaljerad indelning med hänsyn till hur stort behovet av att anpassa antagandet är (givet vilka egenskaper deltagarna har i de program som studeras) och hur kostsamt det är att förändra programmet för effektutvärdering.

1.7 Rapportens disposition

I nästa avsnitt beskrivs tre studier inom Arbetsförmedlingen som analyserar problemet med avaktualisering utan orsak. Data beskrivs i avsnitt tre medan avsnitt fyra inkluderar en beskrivning av de olika metoder som används i den kommande analysen. I avsnitt fem analyseras hur varje egenskap per se påverkar sysselsättningsstatus bland personer som avaktualiserats av okänd orsak. De följande två avsnitten presenterar resultat av att använda de två beskrivna indelningsmetoderna: metoden med ett valt gränsvärde för hur små skillnader kan

vara i avsnitt sex och den statistiska metoden i avsnitt sju. Avsnitt åtta presenterar slutsatser av analysen.

Analysen i rapporten baseras på årsdata. År 2019 blev kontant bruttolön för anställning eller inkomst av näringsverksamhet per månad för första gången tillgänglig i registerdata. Vi kompletterade analysen med olika bedömningar av sysselsättningsstatus de första två månaderna efter avaktualiseringen baserade på månadsdata i bilaga tre. Men det är för tidigt att använda dessa bedömningar i praktiken, flera år med månadsdata behövs för att minska osäkerheten i våra resultat.

2 Tidigare studier

Problemet med avaktualisering av okänd orsak har studerats i tre tidigare rapporter inom Arbetsförmedlingen. Arbetsförmedlingen (2010) redovisar resultat från en urvalsundersökning från 2005 och 2006 riktad till arbetssökande som lämnat Arbetsförmedlingen av okänd orsak och presenterar en imputeringsmodell för att prediktera om individerna har fått arbete eller inte.

Arbetsförmedlingen (2012) studerar sysselsättningsgraden bland de personer som ingår i de utvärderingar som presenteras i Arbetsmarknadsrapporten och har avaktualiserats av okänd orsak. Rapporten undersöker hur stor andel som är sysselsatta enligt SCB:s register ungefär ett år efter programstart. I utvärderingen av arbetsmarknadsutbildning är det ungefär 50 procent som är sysselsatta bland personer som har avaktualiserats av okänd orsak. I utvärderingen av arbetspraktik är andelen drygt 40 procent och i utvärderingen av start av näringsverksamhet ungefär 60 procent.

Med hjälp av en regressionsmodell kan man skatta sannolikheten för varje individ att ha fått ett arbete beroende på individens individuella egenskaper, en så kallad imputeringsmodell.² I Nilsson (2016)³ skattas en imputeringsmodell baserad på enkätdata från 2005/2006 och på fullständiga administrativa data från 2005/2006 och 2011/2012. Modellerna utvärderas i termer av deras förmåga att göra korrekta prediktioner och jämförs med en slumpmässig imputering där slumpen avgör vilka av de arbetssökande som ges utfallet arbete. Imputeringsmodellerna visar sig ha en högre prediktionsförmåga (andelen korrekta prediktioner) än slumpmässig imputering. Imputeringsmodellerna visar till exempel att sysselsättningsnivån bland personer som avregistreras av okänd orsak är lägre för äldre personer, för lågutbildade, för personer födda utanför Europa och för personer med en funktionsnedsättning som medför nedsatt arbetsförmåga.⁴

² Imputering kräver kunskaper om sysselsättningsgraden i gruppen för att kunna välja vilka prediktioner som ska klassificeras som att individen har fått arbete eller inte fått arbete.

³ För en version på svenska se Arbetsförmedlingen (2017).

⁴ Försättningsvis används "funktionsnedsättning" för "funktionsnedsättning som medför nedsatt arbetsförmåga".

3 Data

Det finns två alternativa mått som kan användas för att analysera om en person har fått ett arbete efter avslutat program, ett mått är tillgängligt i Arbetsförmedlingens register och ett är tillgängligt i SCB:s register. Man kan använda båda måtten som utfall i effektutvärderingar.

Måttet som är tillgängligt i Arbetsförmedlingens register (Datalagret) används idag i flera utvärderingar, till exempel i utvärderingen av arbetsmarknadspolitiska program. Fördelen är att utfallet registreras direkt och därför är tillgängligt mycket snabbare än måttet i SCB:s register. Nackdelen är bortfallet som uppstår när arbetssökande lämnar Arbetsförmedlingen utan att informera om de har fått ett arbete. I SCB:s registerdata (databasen LISA) brukar måttet för ett referensår bli tillgängligt ca 16 månader efter referensårets slut. Därför kan program endast utvärderas med SCB:s mått långt efter att programmen avslutats.

En annan nackdel med databasen LISA är att det endast inkluderar årlig information om sysselsättning.⁵ I årliga registerdata finns en variabel som kallas för sysselsättningsstatus (november månad)⁶ där individer klassificeras som sysselsatta om de kan antas ha arbetat minst fyra timmar under november.⁷

Sysselsättningsstatus i databasen LISA svarar inte direkt på frågan om hur många personer som övergått till arbete ett visst antal månader efter avslutat program utan hur många som uppskattas ha haft ett arbete i november varje år.

Avaktualiseringsorsaken och datum för avaktualisering har hämtats från tabellen Insper i Arbetsförmedlingens Datalager. Data från Datalagret har matchats med flera variabler från databasen LISA, inklusive sysselsättningsstatus i november månad och flera individuella egenskaper såsom kön, ålder och födelseland. Från LISA 2019 har även information om kontant bruttolön för anställning eller inkomst av näringsverksamhet per månad hämtats för att analysera övergången till arbete en viss tid efter ett program (i bilaga tre).

Tabell 1 visar antalet personer som har lämnat Arbetsförmedlingen av okänd orsak varje år mellan 2015 och 2019 och vilken andel de representerar av alla de personer som lämnade Arbetsförmedlingen.

⁵ Från och med 2019 finns kontant bruttolön för anställning eller inkomst av näringsverksamhet per månad för första gången tillgängligt i registerdata.

⁶ Variabeln heter SyssStat11 i registerdata för åren 2015 – 2018 och SyssStat19 för 2019.

⁷ Vad som avses med sysselsättning i SCB:s register följer i möjligaste mån Internationella arbetsorganisationens definition, vilken innebär att ett inkomstbringande arbete minst en timme under en mätvecka ska betraktas som förvärsarbete. SCB använder operationella metoder och regler för att genomföra individuella avgränsningar huruvida personen förvärsarbetat minst en timme per vecka eller inte.

Tabell 1: Antal och andel personer som avaktualiserades av okänd orsak per år

	2015	2016	2017	2018	2019	2015–2019
Antal personer	69 125	74 092	75 491	88 443	100 281	81 486
Andel personer	0,19	0,21	0,23	0,26	0,30	0,24

Antalet personer som avaktualiserades av okänd orsak har ökat mellan 2015 och 2019. Även andelen med okänd avaktualiseringsorsak har ökat över tid och var i genomsnitt 24 procent under perioden 2015 – 2019. Vi analyserar inte varför andelen som avaktualiseras av okänd orsak ökar men vi visar att ökningen sker oavsett kön, ålder, utbildningsnivå, inrikes/utrikes födda och eventuell funktionsnedsättning.

Nilsson (2016) identifierar flera individuella egenskaper som kännetecknar dem som har fått jobb efter att ha lämnat Arbetsförmedlingen av okänd orsak. Vår analys inkluderar flera av dessa egenskaper. Tabell 2 och 3 visar dessa egenskaper bland gruppen som har lämnat Arbetsförmedlingen av okänd orsak mellan 2015 och 2019.

Tabell 2 visar hur de personer som avaktualiserades av okänd orsak kan delas upp i kvinnor och män, unga och vuxna, olika utbildningsnivåer, olika födelseland och uppgift om eventuell funktionsnedsättning. Den första raden visar till exempel att 43 procent av de personer som avaktualiserades av okänd orsak mellan 2015 och 2019 var kvinnor, medan resten (57 procent) var män. Andelen unga (under 25 år) och andelen personer med uppgift om eventuell funktionsnedsättning har minskat mellan 2015 och 2019, medan andelen kvinnor och utrikes födda har ökat.

Tabell 2: Andelar av de personer som avaktualiserades av okänd orsak som karakteriseras av olika individegenskapen per år

	2015	2016	2017	2018	2019	2015–2019
Kvinna	0,41	0,43	0,43	0,44	0,44	0,43
Unga (under 25 år)	0,37	0,33	0,29	0,28	0,29	0,31
Förgymnasial utb.	0,33	0,30	0,31	0,30	0,30	0,31
Gymnasial utb.	0,54	0,54	0,50	0,49	0,49	0,51
Eftergymnasial utb.	0,13	0,15	0,16	0,16	0,16	0,15
Okänd utbildning	0,00	0,02	0,03	0,04	0,05	0,03
Inrikes född	0,61	0,58	0,56	0,50	0,53	0,55
Född i EU28	0,06	0,06	0,06	0,06	0,06	0,06
Född utanför EU28	0,33	0,35	0,38	0,36	0,41	0,37
Kod för funkt.	0,11	0,09	0,09	0,08	0,08	0,09

Not: Utb.: utbildning; Kod för funkt.: kod för funktionsnedsättning; EU28: EU (28 länder).
Källa: Arbetsförmedlingens Datalager och SCB:s databas LISA.

Tabell 3 visar hur stor andel av de personer som avaktualiserades varje år som har en viss egenskap. Den första raden visar till exempel att 22 procent av de kvinnor som avaktualiserades mellan 2015 och 2019 lämnade Arbetsförmedlingen utan att berätta varför.

Trots en minskande andel (vilket visas i tabell 2), har unga varit överrepresenterade i gruppen som avaktualiserades av okänd orsak under hela perioden. Bland unga avaktualiserade var det 31 procent som hade okänd orsak under perioden 2015 – 2019, medan motsvarande andel av vuxna var 22 procent. En högre andel män som avaktualiserades hade okänd orsak jämfört med andelen kvinnor. En högre andel personer med okänd eller förgymnasial utbildning som avaktualiserades hade okänd orsak jämfört med personer med gymnasial eller eftergymnasial utbildning. Också en högre andel utrikes födda som avaktualiserades hade okänd orsak jämfört med andelen inrikes födda. Personer med uppgift om funktionsnedsättning avaktualiserades av okänd orsak i mindre utsträckning än övriga (19 procent jämfört med 24 procent). Notera att den ökande trenden som visades i tabell 1 gäller alla grupper. Oavsett egenskaper har andelen avaktualiserade av okänd orsak av de som avaktualiserades ökat mellan 2015 och 2019.

Tabell 3: Andelar av de personer som karakteriseras av olika individegenskapen som avaktualiserades av okänd orsak per år (andel inom gruppen)

	2015	2016	2017	2018	2019	2015–2019
Kvinna	0,17	0,19	0,22	0,24	0,27	0,22
Man	0,21	0,23	0,25	0,28	0,32	0,26
Unga (under 25 år)	0,25	0,28	0,29	0,34	0,42	0,31
Vuxna	0,17	0,19	0,21	0,24	0,26	0,22
Förgymnasial utb.	0,28	0,29	0,32	0,34	0,38	0,32
Gymnasial utb.	0,17	0,20	0,21	0,25	0,29	0,22
Eftergymnasial utb.	0,13	0,15	0,17	0,19	0,21	0,17
Okänd utbildning	0,00	0,41	0,42	0,42	0,46	0,43
Inrikes född	0,16	0,18	0,20	0,23	0,27	0,21
Utrikes född	0,25	0,27	0,29	0,36	0,33	0,31
Född i EU28	0,20	0,23	0,25	0,27	0,30	0,25
Född utanför EU28	0,27	0,28	0,29	0,31	0,34	0,30
Kod för funkt.	0,17	0,17	0,19	0,20	0,23	0,19
Ej kod för funkt.	0,19	0,22	0,24	0,27	0,30	0,24

Not: Utb.: utbildning; Kod för funkt.: kod för funktionsnedsättning; EU28: EU (28 länder).
Källa: Arbetsförmedlingens Datalager och SCB:s databas LISA.

De flesta egenskaper klassificeras på samma sätt i Arbetsförmedlingens Datalager som i databasen LISA, men vi ser stora skillnader i hur utbildningsnivå klassificeras i de två källorna. Vid en jämförelse av utbildningsnivå för personer som avaktualiserades av okänd orsak under perioden 2015 – 2018 kan man se att många personer som enligt LISA-databasen har en eftergymnasial utbildning, har i stället en gymnasieutbildning som högsta avslutade utbildning enligt Datalagret.

Databasen LISA innehåller data från Utbildningsregistret, Datalagret registrerar den utbildningsnivå som de arbetssökande uppger (uppgifterna kontrolleras vid behov). A priori kan man tänka att Utbildningsregistret är en bättre källa för information om utbildningsnivå. Men uppdateringen av systemet för klassificering av utbildning är problematisk och gör det svårare att använda databasen LISA som källa.

Systemet för klassificering av utbildning uppdaterades 2020 och utbildningsnivån i databasen LISA för år 2019 har klassificerats enligt det nya systemet. Bilaga två beskriver viktiga förändringar och effekter av uppdateringen. En effekt är till exempel att ett stort antal personer som avaktualiserades av okänd orsak 2019 klassificerades som gymnasial utbildade efter att ha varit klassificerade som eftergymnasial utbildade i det gamla systemet.

Klassificeringen av utbildning i Datalagret stämmer bättre överens med det uppdaterade systemet för klassificering av utbildning än vad klassificeringen i databasen LISA (upp till 2018) gör. Vi väljer därför att använda Arbetsförmedlingens Datalager som källa för utbildningsnivå. Datalagret har dessutom fördelen att vara den källa som är tillgänglig när man gör utvärderingar. Nackdelen är en något högre andel personer med okänd utbildningsnivå de sista åren.⁸

Tabell 4 visar den genomsnittliga sysselsättningsgraden för alla de personer som har lämnat Arbetsförmedlingen av okänd orsak varje år.

Tabell 4: Sysselsättningsgrad hos personer som avaktualiserades av okänd orsak. Siffrorna är avrundade till två decimaler.

	2015	2016	2017	2018	2019	2015 – 2019
Sysselsättningsgrad	0,41	0,46	0,48	0,47	0,49	0,47

Källa: Arbetsförmedlingens Datalager och SCB:s databas LISA.

Vi kommer nu att studera hur sysselsättningsgraden varierar med de olika egenskaperna. Om man kombinerar två alternativ för kön, tre för utbildningsnivå (högst förgymnasial utbildning, gymnasial utbildning och eftergymnasial utbildning), två för ålder (unga upp till 24 år och vuxna 25 år eller äldre), två för födelseland (inrikes född och utrikes född⁹) och två för uppgift om eventuell funktionsnedsättning (att ha någon kod som indikerar funktionsnedsättning eller inte) blir det 144 delgrupper. Det skulle vara alldeles för många delgrupper, så vi behöver en process för att endast välja de mest relevanta delgrupperna. Delgrupper väljs så att alla individer tillhör en och bara en delgrupp.

Antalet delgrupper behöver vara få av två skäl. Det första skälet är att antalet personer inom varje delgrupp behöver vara tillräckligt många för att skattningarna för varje delgrupp ska bli säkra. Delgrupper med få observationer riskerar att få osäkra skattningar. Det andra skälet är att få delgrupper innebär en mindre belastning än fler delgrupper i framtagandet av de programscript som används för utvärdering.

4 Metoder för att välja delgrupper

Vi söker en metod för att dela de personer som lämnat Arbetsförmedlingen av okänd orsak in i delgrupper utifrån de egenskaper som har störst betydelse för deras jobbchanser. De variabler som påverkar andelen som har gått till arbete är de olika individegenskaper som dessa personer har såsom utbildningsnivå eller ålder och kallas för oberoende variabler. Utfallsvariabeln är andelen i arbete eller

⁸ Att använda Arbetsförmedlingens Datalager som källa för utbildningsnivå innebär att de personer som avaktualiserades av okänd orsak har i genomsnitt något lägre utbildningsnivå än vad som skulle vara fallet om vi använde LISA som källa under åren 2015 – 2018.

⁹ I analysen delades också utrikes födda upp inomeuropeiskt och utomeuropeiskt födda, men skillnaden i sysselsättningsstatus bland de som avaktualiserats av okänd orsak var liten och inkluderas inte i rapporten.

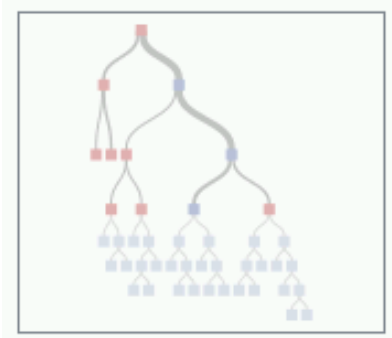
sysselsättningsgrad efter att ha lämnat Arbetsförmedlingen och kallas för beroende variabel.

Processen för att välja delgrupperna görs i flera steg. I avsnitt fem analyseras hur olika egenskaper samvarierar med sysselsättningsstatus bland personer som avaktualiserats av okänd orsak. Två metoder används sedan för att dela in personer som avaktualiserades av okänd orsak i delgrupper. I avsnitt sex väljs delgrupper med hjälp av den statistiska metoden CART (Classification and Regression Tree). I avsnitt sju används ett valt gränsvärde.

Den statistiska metoden CART löser klassificeringsproblem med hjälp av ett beslutsträd. Klassificeringens uppgift är att hitta vilken grupp en individ eller enhet tillhör. Individer delas i klasser som är uttömmande och som ömsesidigt utesluter varandra. Målet är att andelen som gått till arbete i respektive grupp ska vara så lika som möjligt inom varje grupp, och så skilda från varandra som möjligt, mellan grupperna.

Figur 1 visar ett exempel på ett beslutsträd som börjar med en första nod högst upp (som innehåller all data) och sedan förgrenar sig flera gånger. Ett beslutsträd växer fram genom att de oberoende variablerna bedöms i jakt på den specifika skärning som uppfyller ett villkor om maximal inomgruppslikhet och mellangruppskillnad. Därefter upprepas processen stegvis, till något stoppvillkor uppfyllts (eller tills alla grupper är helt och hållet inbördes homogena). CART-algoritmen behöver kriterier för att bestämma hur trädet växer (hur man delar i delgrupper i varje nod) och hur man stoppar trädets tillväxt eller beskär trädet för att säkerställa att det blir lagom stort.

Figur 1: Exempel av beslutsträd



Processen liknar klustring, en annan statistisk metod för att dela in individer eller enheter i klasser. Inom klustring syftar indelningen dock till att maximera likheten utifrån bakgrundsfaktorerna (de oberoende variablerna), medan CART används när syftet är att maximera likhet avseende utfallsvariabeln. Uppgiften liknar även regression på så sätt att det handlar om en förutsägelse baserat på bakgrundsfaktorer, men skillnaden är att istället för att syftet är att förutsägelsen i sig ska komma så nära ett visst värde som möjligt är fokus för klassificeringsuppgiften på indelning i grupper som inbördes har så lika värden av den beroende variabeln som möjligt.

Beslutsträd är bra på att upptäcka interaktioner i data med många olika oberoende variabler. Metoden är icke-parametrisk vilket betyder att den inte gör statistiska antaganden om data. Beslutsträd kan enkelt fånga icke-linjära mönster och är lätta att visualisera och tolka. Dessa fördelar gör att beslutsträd är ett populärt verktyg för att lösa både klassificerings- och regressionsproblem. Danska Star (Danish Agency for Labour Market and Recruitment) har till exempel använt ett beslutsträd i sin profilering av arbetssökande för att identifiera de variabler som är de viktigaste för att förutse risken för långtidsarbetslöshet (Desiere m.fl., 2018).

Beslutsträd har också flera nackdelar. Beslutsträd är känsliga för små variationer i data vilket betyder att en liten ändring i data kan leda till ett mycket annorlunda träd. Denna instabilitet kan leda till felaktiga förutsägelser. En annan nackdel med metoden är att den optimerar ett delningskriterium vid varje nod (en nod i taget), så den hittar lokala optimum som inte behöver vara globala optimum.¹⁰

För att beskriva de kriterier som har används för att bestämma hur trädet växer och slutar växa eller beskärs behöver vi förklara några tekniska begrepp.

Så kallad *entropi* använts som kriterium för att bestämma hur trädet växer. I varje nod, med start i den första noden som innehåller all data, undersöks alla de resterande individegenskaperna (de som inte har använts i tidigare indelningar) för att hitta den variabel som delar in data i två sinsemellan så olika grupper som möjligt med avseende på sysselsättningsgraden. Mer teknisk kan man säga att vid varje nod går algoritmen igenom alla resterande individegenskaper och beräknar så kallad entropi eller informationsvinst för varje potentiell delning av data. Entropi mäter slumpmässighet eller oförutsägbarheten i data. Den minskning av entropi som resulterar av en indelning (för att förgrena noden i två nya barnnoder) kallas för informationsvinst. I analysen väljs individegenskapen med lägst entropi eller högst informationsvinst när noder ska delas så att trädet växer. Det innebär att den individegenskap som har störst inverkan på sysselsättningsgraden används för att förgrena noden i två nya barnnoder. Denna process upprepas på varje barnnod och skapar på så sätt ett träd.

I uppbyggandet av beslutsträdet behöver man också sätta villkor för när det är lämpligt att avsluta varje gren. Om man inte har något villkor alls får man i princip ett träd som är så stort att varje lövnod motsvarar en datapunkt i ursprungsdata. Beskärning är en teknik som minskar storleken på beslutsträd genom att ta bort delar av trädet som är överflödiga. I analysen har vi använt den vanligaste metoden för beskärning, så kallad *cost-complexity pruning* (Breiman m.fl., 1984). Beskärningsmetoden försöker balansera trädets storlek (eller komplexitet) med trädets noggrannhet. I avsnitt sju presenteras tre alternativa beslutsträd med olika storlekar som är resultat av olika val av antal löv under beskärningsprocessen.

¹⁰ En tredje nackdel med beslutsträd är en risk för överanpassning eftersom beslutsträd gör så få antaganden om data de tränas på och har en stor grad av frihet (Géron, 2017). Överanpassning innebär att en modell klarar av att klassificera träningsdata bättre än hur den klarar att klassificera nya data som modellen inte tränats på. Metoden genererar för många nya noder för att passa all data och i slutändan blir trädet för komplext att tolka. Vi har inte fått problem med överanpassning i vår analys.

Algoritmen kan programmeras så att majoriteten av data används vid uppbyggandet av beslutsträdet (så kallad träningsdata) medan resterande data används för att testa felklassificeringar, dvs. trädets noggrannhet. I vårt fall handlar det om andelen personer som klassificeras som icke-sysselsatta trots att de är sysselsatta och tvärtom.

Metoden i avsnitt sju är inte lika automatiserad, i stället används ett valt gränsvärde för att välja delgrupper. Valet startas med sex delgrupper (tre utbildningsnivåer gånger två uppgifter om eventuell funktionsnedsättning) och sedan delas dessa delgrupper upp ytterligare baserat på de tre resterande egenskaperna (ålder, kön och födelseland). Vi kallar resultaten för ”relevanta” delgrupper. Med relevanta delgrupper menas här att dessa delgrupper skiljer sig tillräckligt mycket i sysselsättningsgrad för att det ska vara meningsfullt att ha ett antagande om sysselsättningsgrad per delgrupp i stället för ett generellt antagande för hela gruppen. Nya delgrupper skapas endast om det är en tillräckligt stor skillnad i sysselsättningsstatus mellan dem. Endast skillnader på mer än 12 procent bedöms som relevanta.¹¹ Processen resulterar i 11 relevanta delgrupper.

5 Hur påverkas sysselsättningsgraden av individers egenskaper?

Tidigare studier visar att sysselsättningsnivån bland personer som avregistreras av okänd varierar med ålder, utbildningsnivå, födelseland och funktionsnedsättning som medför nedsatt arbetsförmåga. Som ett första steg för att förstå vilka av dessa egenskaper som borde inkluderas i våra delgrupper studeras i det här avsnittet hur varje egenskap per se påverkar sysselsättningsstatus bland personer som avaktualiserats av okänd orsak. Utbildningsnivån och uppgiften om eventuell funktionsnedsättning identifieras i avsnittet som de mest relevanta egenskaperna som påverkar sysselsättningsgraden.

5.1 Utbildningsnivåer

Skillnader i utbildningsnivå har stor påverkan på sysselsättningsgraden bland personer som avaktualiserats av okänd orsak.

¹¹ Ett alternativt sätt skulle vara att testa om skillnaden mellan olika delgrupper är signifikant skild från noll. Standardavvikelsen är dock väldigt låg i alla beräkningar av sysselsättningsgrad, vilket skulle leda till att vi fick alltför många delgrupper med den metoden.

Tabell 5: Sysselsättningsgrad hos personer som avaktualiserades av okänd orsak per utbildningsnivå. Siffrorna är avrundade till två decimaler.

	2015	2016	2017	2018	2019	2015 – 2019
FG	0,27	0,33	0,34	0,36	0,38	0,34
GY	0,48	0,53	0,54	0,53	0,55	0,53
EG	0,49	0,55	0,56	0,56	0,58	0,55
OU	-	0,28	0,31	0,30	0,32	0,31

Not: FG: förgymnasial utbildning, GY: gymnasial utbildning, EG: eftergymnasial utbildning, OU: okänd utbildning - Källa: AF:s Datalager och SCB:s databas LISA.

Den sista kolumnen i tabell 5 visar att den genomsnittliga sysselsättningsgraden under perioden 2015 – 2019 var ca 56 procent högre för personer med gymnasial utbildning än för personer med förgymnasial utbildning, medan personer med eftergymnasial utbildning har 63 procent högre sysselsättningsgrad än personer med förgymnasial utbildning. Andelen personer med okänd utbildning är väldigt låg (se tabell 1) och därför kommer de att exkluderas i analysen framöver. Exkluderingen gör det lättare att hålla ner antalet delgrupper när flera egenskaper kombineras.

5.2 Uppgift om eventuell funktionsnedsättning

Uppgiften om eventuell funktionsnedsättning samvarierar stort med sysselsättningsgraden. Den sista kolumn i tabell 6 visar att den genomsnittliga sysselsättningsgraden under perioden 2015 – 2019 var 63 procent högre bland personer utan kod för funktionsnedsättning än bland personer med uppgift om eventuell funktionsnedsättning.

Tabell 6: Sysselsättningsgrad uppdelat på om personen har en kod för funktionsnedsättning och övriga. Siffrorna är avrundade till två decimaler.

	2015	2016	2017	2018	2019	2015 – 2019
Kod för funktionsnedsättning	0,24	0,28	0,32	0,32	0,33	0,30
Övriga	0,43	0,48	0,49	0,49	0,51	0,48

Källa: Arbetsförmedlingens Datalager och SCB:s databas LISA.

5.3 Kvinnor och män

Kvinnor har generellt högre sysselsättningsgrad än män (se tabell 7), men skillnaden är inte stor. Sysselsättningsgraden bland kvinnor var endast fem procent högre än bland män.

Tabell 7: Sysselsättningsgrad uppdelat på kön. Siffrorna är avrundade till två decimaler.

	2015	2016	2017	2018	2019	2015 – 2019
Kvinnor	0,43	0,48	0,48	0,48	0,51	0,48
Män	0,40	0,45	0,47	0,47	0,48	0,46

Källa: Arbetsförmedlingens Datalager och SCB:s databas LISA.

5.4 Ålder

Personer under 25 år definieras som unga, medan personer mellan 25 år och 69 år definieras som vuxna. Sysselsättningsgraden bland personer som avaktualiserades av okänd orsak har varierat något med ålder över tid, men skillnaden mellan de ungas och de vuxnas sysselsättningsgrad är nästan obefintlig under hela perioden (se sista kolumn i tabell 8).

Tabell 8: Sysselsättningsgrad uppdelat på ålder. Siffrorna är avrundade till två decimaler.

	2015	2016	2017	2018	2019	2015 – 2019
Unga	0,43	0,47	0,49	0,46	0,47	0,46
Vuxna	0,40	0,45	0,47	0,48	0,50	0,47

Källa: Arbetsförmedlingens Datalager och SCB:s databas LISA.

5.5 Inrikes och utrikes födda

Den sista kolumn i tabell 9 visar att den genomsnittliga sysselsättningsgraden under perioden 2015 – 2019 var 9 procent lägre bland utrikes födda än bland inrikes födda.

Tabell 9: Sysselsättningsgrad uppdelat på ursprung. Siffrorna är avrundade till två decimaler.

	2015	2016	2017	2018	2019	2015 – 2019
Inrikes födda	0,45	0,48	0,50	0,49	0,51	0,49
Utrikes födda	0,36	0,43	0,45	0,45	0,48	0,44
Födda i EU utom Sverige	0,39	0,44	0,46	0,48	0,49	0,46
Utomeuropeiskt födda	0,36	0,42	0,45	0,45	0,48	0,44

Källa: Arbetsförmedlingens Datalager och SCB:s databas LISA.

Eftersom skillnaden i sysselsättningsgrad mellan europeiskt och utomeuropeiskt födda inte är särskilt stor kommer endast inrikes och utrikes födda att inkluderas som viktiga egenskaper i nästföljande analys. Att inte dela upp gruppen utrikes födda ytterligare gör det lättare att hålla nere antalet delgrupper som kombinerar flera egenskaper.

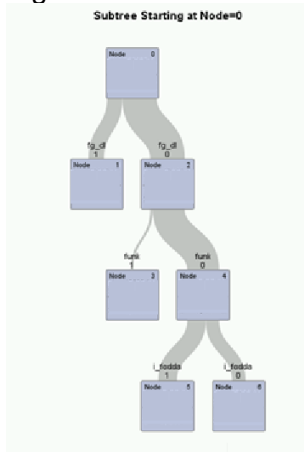
6 Indelning i delgrupper med ett beslutsträd

Den statistiska metoden CART har använts i det här avsnittet för att lösa klassificeringsproblemet med hjälp av ett beslutsträd. Data för 70 procent av de personer som avaktualiserades av okänd orsak (så kallad träningsdata) har använts vid uppbyggandet av beslutsträdet, medan de resterande 30 procent (så kallad valideringsdata) har använts för att testa trädets noggrannhet eller andelen felklassificerade personer. Olika val av maximalt antal löv ger som resultat tre alternativa beslutsträd.

6.1 Indelning i fyra delgrupper

Beslutsträdet ger oss den indelning som anses vara mest effektiv baserad på de kriterier som valts för att bestämma hur trädets växer och hur man stoppar trädets tillväxt eller beskär trädets. Det minsta beslutsträdet är resultatet av att använda cost-complexity pruning¹² utan att specificera ett antal löv. Beslutsträdet visas i figur 2.

Figur 2: Beslutsträd utan specifikation av antal löv



Beslutsträdet delar först personer i två grupper baserat på deras utbildningsnivå (personer med förgymnasial utbildning eller andra utbildningsnivåer). Personer med förgymnasial utbildning delas inte vidare utan blir en så kallad *slutnod*. Personer med minst gymnasial utbildning delas beroende på om de har uppgift om eventuell funktionsnedsättning eller inte. Personer med minst gymnasial utbildning och uppgift om eventuell funktionsnedsättning blir en ny *slutnod*. Personer med minst gymnasial utbildning och utan uppgift om eventuell funktionsnedsättning delas vidare i de två sista *slutnoder*, en för inrikes födda och en för utrikes födda. Fyra *slutnoder* innebär en indelning i fyra delgrupper.

Algoritmen ger information om andel felklassificerade personer både i träningsdata och i valideringsdata. I båda fall gör modellen rätt klassificering för ca 57 procent av de sysselsatta personerna och ca 63 procent av de icke-sysselsatta. Det betyder att

¹² Cost-complexity pruning är en beskärningsmetod som försöker balansera trädets storlek (eller komplexitet) med andelen personer som klassificeras fel (eller trädets noggrannhet).

modellen gör felaktig klassificering för ca 43 procent av de sysselsatta personerna som klassificeras som icke-sysselsatta och för ca 37 procent av de icke-sysselsatta personer som felaktigt klassificeras som sysselsatta. Totalt klassificeras ca 60 procent av personerna rätt av modellen, vilket innebär att noggrannheten inte är speciellt hög. Som jämförelse har modeller med slumpmässig imputering i Nilsson (2016), till exempel, mellan 50 och 53 procent korrekta prediktioner.

I vår analys tar vi den indelning som beslutsträdet föreslår och uppskattar sedan sysselsättningsgraden med hela datasetet, inte bara träningsdata. Tabell 10 visar sysselsättningsgraden när alla personer inkluderas i varje delgrupp.

Tabell 10: Antal personer, antal sysselsatta personer och sysselsättningsgrad i varje delgrupp under perioden 2015 – 2019.

	FG		Ej FG	
		Kod för funkt.	Ej kod för funkt.	
			IF	UF
Antal	124 612	19 816	146 286	103 359
Sysselsatta	42 238	6 867	84 797	52 193
Andel	0,34	0,35	0,58	0,50

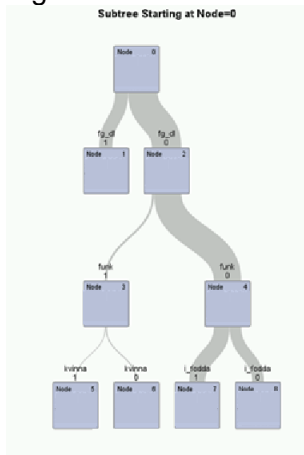
Not: FG: förgymnasial utbildning; Ej FG: övriga utbildningsnivåer; Kod för funkt.: kod för funktionsnedsättning som medför nedsatt arbetsförmåga; IF: inrikes födda; UF: utrikes födda.

Under åren 2015 – 2019 hade personer som avaktualiserades av okänd orsak med högst förgymnasial utbildning en sysselsättningsgrad på 34 procent. Personer som avaktualiserades av okänd orsak med minst gymnasial utbildning och uppgift om eventuell funktionsnedsättning hade en sysselsättningsgrad på 35 procent. Personer som avaktualiserades av okänd orsak med minst gymnasial utbildning och utan uppgift om eventuell funktionsnedsättning hade en sysselsättningsgrad på 58 procent om de var inrikes födda och 50 procent om de var utrikes födda. Dessa uppskattade sysselsättningsgrader kan användas som antagande om sannolikheten att få jobb i framtida utvärderingsprogram.

6.2 Indelning i fem delgrupper

Om vi kompletterar cost-complexity pruning med ett val av ett högre antal löv kan vi få som resultat ett större beslutsträd (se figur 3).

Figur 3: Beslutsträd med högre antal löv



Figur 3 visar ett träd med fem löv. Trädet liknar det i avsnitt 7.1 förutom att även personer med minst gymnasial utbildning och uppgift om eventuell funktionsnedsättning som medför nedsatt arbetsförmåga delas i två delgrupper beroende på kön. Noggrannheten i modellen är väldigt lik den i modellen i avsnitt 7.1, med i princip samma andel felklassificerade personer både för träningsdata, valideringsdata och totalt.

Tabell 11 visar sysselsättningsgraden när alla personer inkluderas i varje delgrupp.

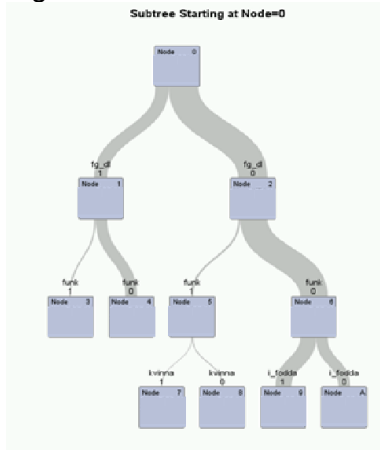
Tabell 11: Antal personer, antal sysselsatta personer och sysselsättningsgrad i varje delgrupp under perioden 2015 – 2019.

	FG		Ej FG		
	Kod för funkt.		Ej kod för funkt.		
	Kvinnor	Män	IF	UF	
Antal	124 612	7 948	11 368	146 286	103 359
Sysselsatta	42 238	3 019	3 741	84 797	52 193
Andel	0,34	0,38	0,33	0,58	0,50

Not: FG: förgymnasial utbildning; Ej FG: övriga utbildningsnivåer; Kod för funkt.: kod för funktionsnedsättning som medför nedsatt arbetsförmåga; IF: inrikes födda; UF: utrikes födda.

6.3 Indelning i sex delgrupper

Figur 4: Beslutsträd med ännu högre antal löv



Figur 4 visar ett träd med sex löv som är ett resultat av cost-complexity pruning med val av ett ännu större antal löv. Trädet liknar det i avsnitt 7.2 förutom att även personer med förgymnasial utbildning delas beroende på om de har uppgift om eventuell funktionsnedsättning som medför nedsatt arbetsförmåga eller inte. Även den här modellen klassificerar nästan precis lika bra som modellen med fyra delgrupper både för träningsdata, valideringsdata och totalt.

Tabell 12 visar sysselsättningsgraden när alla personer inkluderas i varje delgrupp.

Tabell 12: Antal personer, antal sysselsatta personer och sysselsättningsgrad i varje delgrupp under perioden 2015 – 2019.

	FG		Ej FG			
	Kod för funkt.	Ej kod för funkt.	Kod för funkt.		Ej kod för funkt.	
			Kvinnor	Män	IF	UF
Antal	15 946	108 666	7 948	11 368	146 286	103 359
Sysselsatta	3 727	38 511	3 019	3 741	84 797	52 193
Andel	0,23	0,35	0,38	0,33	0,58	0,50

Not: FG: förgymnasial utbildning; Ej FG: övriga utbildningsnivåer; Kod för funkt.: kod för funktionsnedsättning som medför nedsatt arbetsförmåga; IF: inrikes födda; UF: utrikes födda.

6.4 Resultat av beslutsträdsanalysen

Analysen i det här avsnittet indikerar att få delgrupper kan vara tillräckligt för att förbättra det tidigare antagandet om att 50 procent av de som avaktualiseras av okänd orsak har fått ett arbete. De tre alternativa beslutsträden i avsnitt 6.1, 6.2 och 6.3 klassificerar personer som sysselsatta eller icke-sysselsatta i princip lika bra. Men noggrannheten i analysen är inte särskild hög, cirka 40 procent av personerna klassificeras fel i de tre alternativa indelningarna. Detta kan till exempel vara ett

resultat av datasetets storlek, fler observationer skulle kunna öka noggrannheten. Datasetets storlek kan också vara ett problem med tanke på att beslutsträd är känsliga för små variationer i data vilket betyder att en liten ändring i data kan leda till ett mycket annorlunda träd. Flera observationer skulle kunna minska känsligheten. Ett sätt att öka antal observationer skulle vara att inkludera tidigare år i analysen. Samtidigt skulle data då riskera vara sämre anpassad till dagens arbetssökande. Detta är speciellt problematiskt nu eftersom det har varit stora förändringar i de arbetssökandes sammansättning över de sista åren.

Automatiska metoder är svåra att tränga igenom. Vi kompletterar analysen med en mer manuell metod där vi i stället väljer ett gränsvärde för skillnad i sysselsättningsgrad.

7 Indelning i delgrupper med ett gränsvärde för skillnad i sysselsättningsgrad

Utbildningsnivån och uppgiften om eventuell funktionsnedsättning identifierades i avsnitt fem som de mest relevanta egenskaperna som påverkar sysselsättningsgraden. Det här avsnittet analyserar om dessa delgrupper behöver delas upp ytterligare baserat på de tre resterande egenskaperna med hjälp av ett gränsvärde för skillnad i sysselsättningsgrad.

Endast uppdelningar där det är en tillräckligt stor skillnad i sysselsättningsgrad mellan de nya delgrupperna behålls. Vi vill ha en relativt stor skillnad mellan olika delgrupper och bedömer därför att endast skillnader på mer än 12 procent är relevanta. Den exakta gränsen valdes med hänsyn till faktiska skillnader i sysselsättningsgrad mellan de olika grupper i vår data. Mindre skillnader bedömdes resultera i för många grupper som skulle belasta framtida utvärderingsprogram samtidigt som belastningen med en gräns av 12 procent var inte så stor för att behöva välja ännu större skillnader.

7.1 Kombinationer av utbildningsnivå och kod för funktionsnedsättning

Tabell 13 visar kombinationer av de viktigaste egenskaperna som identifierades i avsnitt 4. I sista kolumnen kan vi konstatera att dessa egenskaper fortfarande är viktiga även efter att ha kombinerats. Bland personer med uppgift om eventuell funktionsnedsättning är sysselsättningsgraden 47 procent högre om de har gymnasial utbildning i stället för förgymnasial utbildning och 15 procent högre om de har eftergymnasial utbildning i stället för gymnasial utbildning. Bland övriga är sysselsättningsgraden 54 procent högre om de har gymnasial utbildning i stället för förgymnasial utbildning, men endast 3 procent högre om de har eftergymnasial utbildning i stället för gymnasial utbildning. Bland personer med förgymnasialutbildning / gymnasial / eftergymnasial utbildning är

sysselsättningsgraden 34 / 37 / 30 procent lägre om de har en kod för funktionsnedsättning än om de inte har en sådan kod. De flesta procentuella skillnaderna är högre än vårt valda gränsvärde på 12 procent.¹³

Tabell 13: Sysselsättningsgrad uppdelat på om personen har en kod för funktionsnedsättning och utbildningsnivå. Siffrorna är avrundade till två decimaler.

		2015	2016	2017	2018	2019	2015 – 2019
Kod för funkt.	FG	0,17	0,23	0,24	0,27	0,27	0,23
	GY	0,29	0,32	0,37	0,34	0,38	0,34
	EG	0,32	0,36	0,40	0,44	0,42	0,39
Övriga	FG	0,28	0,34	0,36	0,37	0,40	0,35
	GY	0,50	0,55	0,56	0,55	0,56	0,54
	EG	0,50	0,56	0,57	0,56	0,59	0,56

Not: Kod för funkt.: kod för funktionsnedsättning; FG: förgymnasial utbildning; GY: gymnasial utbildning; EG: eftergymnasial utbildning.

Källa: Arbetsförmedlingens Datalager och SCB:s databas LISA.

Nästa steg är att se om man borde dela dessa sex delgrupper ytterligare med hänsyn till resterande egenskaper.

7.2 Indelning av kombinationer av utbildningsnivå och uppgift om eventuell funktionsnedsättning

Den tredje och fjärde kolumnen i tabell 14 visar de inrikes (IF) och utrikes (UF) föddas sysselsättningsgrad under perioden 2015 – 2019, medan den procentuella skillnaden mellan deras sysselsättningsgrad visas i den femte kolumnen. Kolumnerna 6 – 8 jämför på samma sätt unga (under 25 år) och vuxna (25 år och äldre) och kolumnerna 9 – 11 jämför på samma sätt kvinnor och män under samma period.

¹³ När man väl börjar få många delgrupper blir antal personer i varje grupp en viktig fråga eftersom det påverkar osäkerheten i den uppskattade sysselsättningsgraden. Delgrupper av personer med kod för funktionsnedsättning är klart mindre än övriga. Det lägsta antal personer i en delgrupp i tabell 8 är 1 839 bland personer med gymnasial utbildning och kod för funktionsnedsättning år 2017. Delgruppen har en standardavvikelse av 1,11 procent, den största bland alla delgrupper i tabell 7. I sista kolumnen med genomsnittet 2015 – 2019 är det lägsta antalet 8 323 bland personer med gymnasial utbildning och kod för funktionsnedsättning.

Tabell 14: Sysselsättningsgrad under perioden 2015 – 2019 uppdelat på om personen har uppgift om eventuell funktionsnedsättning och utbildningsnivå kombinerat med övriga egenskaper. Siffrorna är avrundade till två decimaler.

		IF	UF	% sk.	Unga	Vuxna	% sk.	Kvinnor	Män	% sk.
Kod för funkt.	FG	0,24	0,22	0,07	0,23	0,23	0,00	0,25	0,23	0,10
	GY	0,35	0,32	0,10	0,38	0,34	0,11	0,37	0,33	0,12
	EG	0,42	0,35	0,18	0,23	0,39	-0,42	0,44	0,34	0,31
Övriga	FG	0,34	0,37	-0,07	0,32	0,37	-0,15	0,34	0,36	-0,06
	GY	0,57	0,51	0,12	0,57	0,53	0,06	0,57	0,53	0,08
	EG	0,65	0,50	0,30	0,55	0,56	-0,03	0,55	0,57	-0,03

Not: Kod för funkt.: kod för funktionsnedsättning som medför nedsatt arbetsförmåga; IF: inrikes födda; UF: utrikes födda; FG: förgymnasial utbildning; GY: gymnasial utbildning; EG: eftergymnasial utbildning. Standardavvikelsen är låg för samtliga mått. % sk: procentuell skillnad

Källa: Arbetsförmedlingens Datalager och SCB:s databas LISA.

Bland personer med uppgift om eventuell funktionsnedsättning och förgymnasial eller gymnasial utbildning finns ingen skillnad högre än det valda gränsvärdet på 12 procent. Vi bedömer därför att skillnaderna inte är tillräckligt stora för att vara relevanta i analysen, därför väljer vi att inte dela de här delgrupperna ytterligare.

Bland personer med uppgift om eventuell funktionsnedsättning och eftergymnasial utbildning är skillnaderna baserade på alla tre egenskaper relevanta (högre än det valda gränsvärdet 12 procent). Samtidigt finns det endast 22 eftergymnasialt utbildade unga med uppgift om eventuell funktionsnedsättning.¹⁴ Vi bedömer att det är bättre att inte dela upp delgruppen ytterligare på ålder på grund av gruppens ringa storlek. Vi delar gruppen i fyra delgrupper baserade på födelseland och kön.

Bland övriga personer med förgymnasial utbildning är skillnaden i sysselsättningsgrad baserat på ålder relevant, medan bland övriga personer med eftergymnasial utbildning är endast skillnaden baserad på födelseland relevant. För övriga personer med gymnasial utbildning finns ingen skillnad som överstiger det valda gränsvärdet på 12 procent. Vi bedömer därför att skillnaderna inte är tillräckligt stora för att vara relevanta i analysen och väljer att inte dela den här delgruppen ytterligare. Bland övriga personer med eftergymnasial utbildning är endast skillnaden baserad på födelseland relevant. Totalt blir det 11 olika delgrupper.

7.3 Resultat av analysen med hjälp av ett gränsvärde för skillnad i sysselsättningsgrad

Tabell 15 visar antal personer, antal sysselsatta personer och sysselsättningsgrad inom varje av de 11 relevanta delgrupper. I tabell 15a ingår endast personer med

¹⁴ Av personer avaktualiserade av okänd orsak med kod för funktionsnedsättning och eftergymnasial utbildning får 22 unga, 2770 vuxna, 1798 inrikes födda, 994 utrikes födda, 1445 kvinnor och 1347 män.

uppgift om eventuell funktionsnedsättning medan i tabell 15b ingår personer utan uppgift om eventuell funktionsnedsättning.

Tabell 15a: Antal personer, antal sysselsatta personer och sysselsättningsgrad i varje relevant delgrupp under perioden 2015 – 2019. Personer med uppgift om eventuell funktionsnedsättning.

	FG		GY		EG			
					IF		UF	
			Kvinnor	Män	Kvinnor	Män		
Antal	15 946	16 524	1001	797	444	550		
Sysselsatta	3 727	5 663	470	278	171	178		
Andel	0,23	0,34	0,47	0,35	0,39	0,32		

Tabell 15b: Antal personer, antal sysselsatta personer och sysselsättningsgrad i varje relevant delgrupp under perioden 2015 – 2019. Personer utan uppgift om eventuell funktionsnedsättning.

	FG		GY		EG	
	Unga	Vuxna			IF	UF
Antal	38 762	69 904	189 873	23 447	36 325	
Sysselsatta	12 353	26 158	103 473	15 285	18 232	
Andel	0,32	0,37	0,54	0,65	0,50	

Not: FG: förgymnasial utbildning; GY: gymnasial utbildning; EG: eftergymnasial utbildning; IF: inrikes födda; UF: utrikes födda. Källa: Arbetsförmedlingens Datalager och SCB:s databas LISA.

När vi använder ett gränsvärde för skillnad i sysselsättningsgrad får vi som resultat en indelning i elva grupper, vilket är flera än vad som var fallet med beslutsträd. Vi ser inte detta som ett problem, utan som en möjlighet att välja en mer eller mindre detaljerad indelning. Om behovet att anpassa antagandet (givet vilka egenskaper deltagarna har i de program som studeras) är stor och kostnad med att förändra programmet för effektutvärdering för att ha fler delgrupper är låg så kan effektutvärderingsprogram basera sina antaganden om sannolikheten att ha fått arbete bland dem som lämnat Arbetsförmedlingen av okänd orsak på de uppskattade sysselsättningsgraderna i de 11 relevanta delgrupperna i tabeller 12a och 12b.

8 Slutsats

Resultat av analyser och utvärderingar som använder avaktualiseringsorsak som utfall kan bli snedvridna på grund av bortfallet när avaktualiseringsorsaken är okänd. Vi föreslår ett enkelt sätt att hantera bortfallet. Antagandet om andelen personer som har lämnat Arbetsförmedlingen för att de har fått ett arbete men inte informerat avaktualiseringsorsaken anpassas till egenskaper hos de personer som ingår i analysen eller i deltagar- och jämförelsegrupp i utvärderingen.

Indelningarna i delgrupper har föreslagits som resultat av en analys av sysselsättningsgraden hos de personer som har lämnat Arbetsförmedlingen utan att informera om de har fått jobb eller inte under perioden 2015 – 2019. Målet är att andelen som gått till arbete ska vara så lika som möjligt inom varje delgrupp, och så skilda från varandra som möjligt, mellan delgrupperna. Den statistiska metoden CART ger tre alternativa indelningar med fyra till sex delgrupper. När vi i stället använder ett gränsvärde för skillnad i sysselsättningsgrad får vi som resultat elva delgrupper. Det finns fördelar och nackdelar med båda metoderna. En nackdel med den statistiska metoden är att beslutsträd är känsliga för små variationer i data vilket betyder att en liten ändring i data kan leda till ett mycket annorlunda träd. Nackdelen med den andra metoden är att det valda gränsvärdet är en bedömningsfråga. Ett lägre gränsvärde skulle ha resulterat i fler delgrupper och tvärtom.

De två metoderna ger en möjlighet att välja en mer eller mindre detaljerad indelning. Behovet av att använda en mer detaljerad indelning är större när analysen eller utvärderingen gäller personer som skiljer sig mer från den genomsnittliga personen som avaktualiseras av okänd orsak. Till exempel användes den mer detaljerade indelning som vår metod föreslår i rapporten Arbetsförmedlingen (2021) som fokuserar på unga med funktionsnedsättning som medför nedsatt arbetsförmåga. Det är också viktigare att använda en mer detaljerad indelning när man utvärderar ett program med större skillnader mellan andelen deltagare och andelen jämförelsepersoner som avaktualiseras av okänd orsak. Valet av en mer eller mindre detaljerad indelning behöver även ta hänsyn till programsripten som används för effektutvärdering.¹⁵

En av de effektutvärderingar som tidigare har antagit att de programdeltagare och personer i jämförelsegruppen som lämnat Arbetsförmedlingen av okänd orsak har fått ett arbete med 50 procent sannolikhet är effektutvärderingar av arbetsmarknadspolitiska program inom Arbetsförmedlingen. Antagandet belastade koden väldigt lite och därför kommer vi framöver att anpassa antagandet med hjälp av den mer detaljerade indelningen i elva relevanta delgrupper.

År 2019 blev kontant bruttolön för anställning eller inkomst av näringsverksamhet per månad för första gången tillgängligt i registerdata från SCB. Om några år kan analysen i den här rapporten baseras på månadsdata i stället för årsdata.¹⁶ Men flera år med månadsdata behövs för att minska osäkerheten i analysen.

¹⁵ En mer detaljerad indelning riskerar att belasta scripten med för mycket kod och göra datakörningarna långsammare.

¹⁶ Ett exempel av en sådan analys visas i bilaga 3.

Referenser

- Arbetsförmedlingen (2010), "Arbetssökande som lämnar Arbetsförmedlingen av okänd orsak", Working Paper 2010:1, Arbetsförmedlingen, Stockholm.
- Arbetsförmedlingen (2012), "Avaktualisering av okänd orsak – antaganden vid utvärderingsstudier av program", Working Paper 2012:3, Arbetsförmedlingen, Stockholm.
- Arbetsförmedlingen (2017), "Imputeringsmodell för saknad sysselsättningsstatus i arbetslöshetsdata", Arbetsförmedlingen, Stockholm.
- Arbetsförmedlingen (2021), "Unga med funktionsnedsättning på Arbetsförmedlingen: En fördjupad analys", Arbetsförmedlingen analys 2021:20.
- Breiman, L., J. Friedman, R.A. Olshen och C.J. Stone (1984). "Classification and Regression Trees", Belmont, CA: Wadsworth.
- Desiere, S., K. Langenbucher och L. Struyven (2018), "Statistical Profiling in Public Employment Services: an international comparison", OECD Social, Employment and Migration Working Papers.
- Géron A. (2017), "Hands-On Machine Learning with Scikit-Learn & TensorFlow", O'Reilly Media Inc., USA.
- Nilsson, P. (2016), "An Imputation Model for Dropouts in Unemployment Data", Journal of Official Statistics, Vol. 32, No. 3.

Bilaga 1: Jämförelse av klassificering av utbildning i databasen LISA och i Arbetsförmedlingens Datalagret

Databasen LISA innehåller data från Utbildningsregistret, medan Datalagret registrerar den utbildningsnivå som de arbetssökande uppger (uppgifterna kontrolleras vid behov).

Tabell 16 jämför hur många personer som har klassificerats som förgymnasialt, gymnasialt och eftergymnasialt utbildade i de två tillgängliga källorna. Under perioden 2015 – 2018 har ett stort antal personer (fler än 20 000 varje år) som avaktualiserades av okänd orsak klassificerats som att de har nått eftergymnasial utbildning i databasen LISA men klassificerats som att de endast har nått gymnasial utbildning i Datalagret. År 2019 är antalet personer som har klassificerats som eftergymnasialt utbildade i LISA men gymnasialt utbildade i Datalagret något mindre (5 853 personer). Utbildningsnivåns klassificering i Datalagret stämmer bättre överens med den uppdaterade utbildningsklassifikationen från 2020 än med den tidigare versionen.

Men även under 2019 finns det skillnader i klassificering. Korrelationen mellan förgymnasialt utbildade i LISA och förgymnasialt utbildade i Datalagret är 64 procent. Korrelationen är något lägre (60 procent) för gymnasialt utbildade personer och något högre (73 procent) för eftergymnasialt utbildade personer.

A priori kan man tänka att Utbildningsregistret är en bättre källa för information om utbildningsnivå. Men uppdateringen av systemet för klassificering av utbildning är problematisk och gör det svårare att använda databasen LISA som källa. Vi väljer därför att använda Arbetsförmedlingens Datalager som källa för utbildningsnivå. Datalagret har dessutom fördelen att vara den källa som är tillgänglig när man gör utvärderingar. Nackdelen är en något högre andel personer med okänd utbildningsnivå de sista åren.

Tabell 16: Antal personer per utbildningsnivå år 2015 enligt databasen LISA och Datalagret bland personer som avaktualiserades av okänd orsak år 2019

		Datalagret				
		FG	GY	EG	Ok	
LISA	2015	FG	17 316	3 806	206	0
		GY	3414	9 885	481	0
		EG	1778	23 526	8 225	0
		Ok	248	68	172	0
	2016	FG	17 431	4 048	205	814
		GY	3540	10 673	580	116
		EG	1866	24 326	9 729	92
		Ok	231	74	213	154
	2017	FG	17 317	3 734	203	1672
		GY	3791	10 882	679	279
		EG	2101	22 960	10 763	196
		Ok	219	86	242	367
	2018	FG	19 738	4 349	280	2365
		GY	4311	12 752	758	438
		EG	2474	26 247	13 172	299
		Ok	269	92	277	622
2019	FG	21 871	4 757	253	3365	
	GY	6926	38 179	2138	802	
	EG	533	5 853	13 961	177	
	Ok	328	100	40	998	

Not: FG: förgymnasial utbildning, GY: gymnasial utbildning, EG: eftergymnasial utbildning, Ok: okänd utbildning.

Källa: Arbetsförmedlingens Datalager och SCB:s databas LISA.

Bilaga 2: Uppdateringen av systemet för klassificering av utbildning

Systemet för klassificering av utbildning (Svensk utbildningsklassifikation, SUN) uppdaterades 2020 och utbildningsnivån i databasen LISA för år 2019 har klassificerats enligt det nya systemet. I den här bilagan beskriver vi viktiga förändringar i SUN 2020 och effekter av uppdateringen.

Förändringar i SUN 2020

Några viktiga förändringar är:

- En ny läroplan infördes i gymnasieskolan från och med läsåret 2011/2012. Den nya läroplanen, GY11, innebar förändringar i ämnesplaner, examen, betygsskala, gymnasieprogram och behörighetskrav till gymnasieskolan.
- Högskolan har förändrats i och med Bolognaanpassningen som gäller från den 1 juli 2007. Högskoleutbildningen delas härefter in i grundnivå, avancerad nivå och forskarnivå där varje examen kopplas till en utbildningsnivå med en fast poängomfattning.
- År 2009 infördes yrkeshögskoleutbildningar (YH), en utbildningsform på eftergymnasial nivå.
- Utbildningsformen Kompletterande utbildningar (KU) saknas i den förra versionen av SUN. Utbildningsformen har därefter ersatts med Konst- och kulturutbildningar och vissa andra utbildningar. Utbildningsnomenklaturen ska täcka utbildningar som befolkningen har genomgått och därför beskrivs både Kompletterande utbildningar samt Konst- och kulturutbildningar i SUN 2020. Kompletterande utbildningar och Konst och kulturutbildningar är främst högskoleförberedande konstnärliga utbildningar och yrkesutbildningar inom kulturområdet. I motsats till de kompletterande utbildningarna är samtliga konst- och kulturutbildningar eftergymnasiala.
- Den internationella klassifikationen ISCED har uppdaterats genom ISCED 2011 (nivå) och ISCED-F 2013 (inriktning). Strukturen har förändrats för både nivå- som inriktningsindelningen.

Effekt av uppdateringen i SUN 2020

Med hjälp av registerdata från 2018 kan vi se vilka utbildningsnivåer de personer som lämnade Arbetsförmedlingen år 2019 hade ett år tidigare. Det är exakt samma personer som följs i registerdata bakåt i tiden. En jämförelse av deras utbildningsnivå år 2018 och 2019 hjälper oss att förstå hur uppdateringen har påverkat data.

Tabell 17 visar att de flesta personerna klassificerades som att de hade samma utbildningsnivå år 2018 och 2019. Till exempel hade 28 895 personer förgymnasial utbildning som registrerad utbildningsnivå både 2018 och 2019.

Några personer klassificerades med högre utbildningsnivå år 2019 än 2018, vilket kan vara ett resultat av studier, av uppdateringen av systemet för klassificering av utbildning eller av att klassificeringen hade varit felaktigt. Till exempel klassificerades 3 459 personer som att de hade gymnasial utbildning 2019 efter att ha klassificerats som att de hade förgymnasial utbildning år 2018.

Totalt hade 44 302 av de personer som avaktualiserades av okänd orsak år 2019 klassificerats som att de hade eftergymnasial utbildning år 2018. Av dessa har 19 197 personer fortfarande eftergymnasial utbildning som registrerad utbildningsnivå år 2019. Men fler än hälften (25 060 personer) har klassificerats som gymnasialt utbildade år 2019. Några personer som hade klassificerats som att de hade gymnasial eller eftergymnasial utbildning år 2018 har dessutom klassificerats som förgymnasialt utbildade år 2019. En stor del av dessa förändringar är sannolikt ett resultat av uppdateringen, till exempel av att kompletterande utbildningar inkluderades i systemet som gymnasiala utbildningar. Några förändringar kan också vara resultat av att klassificeringen hade varit felaktigt.

Tabell 17: Antal personer inom gruppen avaktualiserades av okänd orsak år 2019 som hade varje möjlig kombination av registrerad utbildningsnivå i databasen LISA år 2018 och år 2019.

		2019		
		FG	GY	EG
2018	FG	28 895	3 459	70
	GY	200	18 467	171
	EG	45	25 060	19 197

Not: FG: förgymnasial utbildning, GY: gymnasial utbildning, EG: eftergymnasial utbildning.
Källa: Arbetsförmedlingens Datalager och SCB:s databas LISA.

Som ett extra test för att förstå konsekvenserna av uppdateringen år 2019 har jag upprepat samma analys med en jämförelse av registrerad utbildningsnivå år 2017 och år 2018 bland de personer som avaktualiserades av okänd orsak år 2019.

Tabell 18 visar att endast 300 personer (231+59) hade klassificerats som eftergymnasialt utbildade år 2017, som inte nådde samma nivå år 2018. Dessa förändringar är mycket färre än vad som var fallet i tabell 17 och kan vara resultat av att klassificeringen varit felaktigt. Antal personer som hade klassificerats som gymnasialt utbildade år 2018 och förgymnasialt utbildade år 2019 (203) liknar antalet personer i samma cell i tabell 17 och kan också vara resultat av att klassificeringen varit felaktigt. En jämförelse mellan tabell 17 och 18 indikerar att uppdateringen år 2019 har haft störst påverkan på personer som hade klassificerats som om de hade eftergymnasial utbildning tidigare år.

Tabell 18: Utbildningsnivå år 2017 och år 2018 bland personer som avaktualiserades av okänd orsak år 2019

		2018		
		FG	GY	EG
2017	FG	30 589	1 333	3 754
	GY	203	16 088	530
	EG	59	231	37 759

Not: FG: förgymnasial utbildning, GY: gymnasial utbildning, EG: eftergymnasial utbildning.
Källa: Arbetsförmedlingens Datalager och SCB:s databas LISA.

Uppdateringen av systemet för klassificering av utbildning har haft stor påverkan på utbildningsnivån hos personer som lämnat Arbetsförmedlingen. Men sysselsättningsgraden bland personer som har klassificerats som gymnasialt utbildade år 2019 trots att de hade klassificerats som eftergymnasialt utbildade år 2018 (59 procent) är inte lägre än sysselsättningsgraden bland personer som klassificerats som eftergymnasialt utbildade båda åren (58 procent, se tabell 19).

Däremot har de personer som har klassificerats som förgymnasialt utbildade år 2019 trots att de hade klassificerats som gymnasialt utbildade år 2018, klart lägre sysselsättningsgrad (36 procent) än de personer som klassificerats som gymnasialt utbildade båda åren (50 procent). De har samma sysselsättningsgrad som de personer som klassificerats som förgymnasialt utbildade båda åren (36 procent).

Tabell 19: Sysselsättningsgrad år 2019 fördelat på utbildningsnivå år 2018 och år 2019 bland personer som avaktualiserats av okänd orsak år 2019

		2019		
		FG	GY	EG
2018	FG	0,36	0,56	0,57
	GY	0,36	0,50	0,44
	EG	0,31	0,59	0,58

Not: FG: förgymnasial utbildning, GY: gymnasial utbildning, EG: eftergymnasial utbildning.
Källa: Arbetsförmedlingens Datalager och SCB:s databas LISA.

Slutligen analyserar jag vilka de personer är som har klassificerats med gymnasial utbildning år 2019 trots att de hade klassificerats som eftergymnasialt utbildade år 2018. Tre av fyra personer i gruppen är inrikes födda, en av fem är född utanför Europa och resten inom EU (28 länder). Fyra av tio är unga, 43 procent är kvinnor och 6 procent har kod för funktionsnedsättning. Det innebär en överrepresentation av inrikes födda och unga, medan kvinnor och utomeuropeiskt födda är underrepresenterade jämfört med alla de personer som avaktualiserades av okänd orsak år 2019. Det är svårt att förstå varför just dessa personer skulle byta klassificering. Att till exempel ha deltagit i kompletterande utbildningar kan inte direkt relateras till dessa egenskaper.

Bilaga 3: Analys av månadsdata år 2019

Från och med år 2019 inkluderar databasen LISA även information om kontant bruttolön för anställning eller inkomst av näringsverksamhet per månad. Detta innebär att för de personer som lämnade Arbetsförmedlingen år 2019 kan vi bedöma om de var sysselsatta någon av de första två månaderna efter avaktualiseringen. Bedömningen baseras på bruttolönen under dessa två månader.

För att bedöma om en person är sysselsatt behövs ett gränsvärde för bruttolönen. Vi har använt tre olika alternativ i analysen: att ha en positiv inkomst, att ha en inkomst över 8 000 kronor per månad och att ha en inkomst över 15 000 kronor per månad. I tabell 20 beskriver vi hur sysselsättningsgraden förändras om databasen LISA:s variabel för sysselsättningsstatus i november (baserad på årliga registerdata) ersätts av en bedömning av sysselsättningsstatus enligt inkomsten någon av de första två månader efter avaktualiseringen.

Att bedöma alla personer med positiv inkomst som sysselsatta någon av de första två månaderna efter avaktualiseringen ger en högre sysselsättningsstatus än vad som är fallet med LISA:s variabel för sysselsättningsstatus i november. Ju högre inkomst som krävs för att bedöma en person som sysselsatt, desto lägre blir sysselsättningsgraden.

Det är tyvärr inte möjligt att göra den här analysen för tidigare år, så de alternativ som finns kvar är att endast använda månadsdata för 2019. Vår bedömning är att osäkerheten skulle vara för stor och därför behöver en sådan analys vänta ytterligare ett eller två år.

Tabell 20: Sysselsättningsgrad hos personer som avaktualiserades av okänd orsak år 2019 enligt olika definitioner /bedömningar och fördelat på olika egenskaper.

	Antal	Årsdata	Månadsdata		
			Ink>0	Ink>8 000	Ink>15 000
Total	100 281	0,49	0,65	0,55	0,47
Kön					
Kvinna	44 376	0,51	0,65	0,54	0,45
Man	55 905	0,48	0,65	0,56	0,49
Ålder					
Unga	29 227	0,47	0,71	0,56	0,45
Vuxna	71 054	0,50	0,63	0,55	0,48
Utbildning					
FG	29 658	0,38	0,55	0,45	0,37
GY	48 889	0,55	0,72	0,62	0,53
EG	16 392	0,58	0,69	0,61	0,55
OU	5 342	0,32	0,49	0,38	0,31
Ursprung					
Inrikes född	53 535	0,51	0,68	0,57	0,49
Utrikes född	46 746	0,48	0,62	0,53	0,46
Född i EU	5 755	0,49	0,64	0,56	0,47
Utanför EU	40 991	0,48	0,61	0,53	0,46
Funktionsnedsättning					
Kod för funkt.	7 641	0,33	0,46	0,38	0,32
Övriga	92 640	0,51	0,64	0,57	0,49

Not: FG: förgymnasial utbildning; GY: gymnasial utbildning; EG: eftergymnasial utbildning; IF: inrikes födda; UF: utrikes födda. Ink: kontant bruttolön för anställning eller inkomst av näringsverksamhet per månad, Kod för funkt.: kod för funktionsnedsättning som medför nedsatt arbetsförmåga. Källa: Arbetsförmedlingens Datalager och SCB:s databas LISA