

## Indvandrernes medbragte uddannelse

Oplysninger om indvandrernes medbragte uddannelse indgår i statistikken over befolkningens højest fuldført uddannelse, som udarbejdes af Danmarks Statistik.

Statistikken over højest fuldført uddannelse er primært baseret på oplysninger fra administrative registre, der laves og anvendes på danske uddannelsesinstitutioner. Derudover anvendes oplysninger fra en række andre kilder. Det kan fx være oplysninger fra den spørgeskemabaserede Folke- og boligælling i 1970 samt uddannelsesoplysninger indhentet i forbindelse med autorisering eller anden vurdering af en udenlandsk uddannelse foretaget af en dansk myndighed eller en dansk uddannelsesinstitution.

Danmarks Statistik har desuden gennemført en række spørgeskemaundersøgelser af indvandrernes medbragte uddannelse. Disse spørgeskemaundersøgelser har været målrettet personer, hvor der ikke er andre kilder til at fastslå uddannelsesniveaue. For hele befolkningen (15-69 år) i 2017 stammer 7,4 pct. af oplysningerne om højest fuldført uddannelse fra spørgeskemaundersøgelserne om indvandrernes medbragte uddannelse.

De indsamlede uddannelsesoplysninger om indvandrernes medbragte uddannelse suppleret med den foretagne imputation bidrager til at give et mere retvisende billede af den faktiske uddannelsesfordeling i befolkningen og specielt blandt indvandrerne både på nationalt og kommunalt niveau, end det var tilfældet før indsamlingen af disse data.

I dette notat redegøres der for de forskellige undersøgelser, hvornår de er gennemført og hvilket design de er gennemført efter.

## Indhold

<b>Undersøgelse af indvandrernes medbragte uddannelse.....</b>	<b>3</b>
Betydningen af spørgeskemadata for kvaliteten af uddannelsesstatistikken .....	4
<b>Undersøgelser dækkende personer indvandret i årene 1970-2006..</b>	<b>5</b>
Dataindsamling fra sprogcentre.....	5
Dataindsamling hos Danmarks Statistik .....	5
Imputering – korrektion for manglende besvarelser.....	5
<b>Undersøgelse dækkende personer indvandret i årene 2005-2016 ...</b>	<b>6</b>
Dataindsamling på sprogcentre .....	6
Dataindsamling hos Danmarks Statistik .....	8
Databehandling.....	8
Imputation – korrektion af manglende besvarelser .....	8
Validering af imputationen i 2016-undersøgelsen.....	8
<b>Undersøgelse dækkende personer indvandret i årene 2016-2018..</b>	<b>10</b>
Dataindsamling fra sprogcentre.....	10
Dataindsamling hos Danmarks Statistik .....	10
Databehandling.....	10
Imputation – korrektion af manglende besvarelser .....	10
<b>Undersøgelse dækkende personer indvandret i årene 2016-2019...</b>	<b>11</b>
Revision af kvalifikationsregistret.....	11
Populationen.....	11
Dataindsamling fra sprogcentre.....	12
Dataindsamling fra jobcentre .....	12
Dataindsamling hos Danmarks Statistik .....	12
Databehandling.....	12
Imputation – korrektion af manglende besvarelser .....	12
<b>Undersøgelse dækkende personer indvandret i 2019-2020 .....</b>	<b>13</b>
Populationen.....	13
Dataindsamling fra sprogcentre.....	13
Dataindsamling fra jobcentre .....	13
Dataindsamling hos Danmarks Statistik .....	13
Databehandling.....	13
Imputation – korrektion af manglende besvarelser .....	14
<b>Oversigt over undersøgelserne .....</b>	<b>15</b>
Indvandrede efter 1970 og før 1. januar 1999 .....	15
Indvandrede i 1999-2000 .....	15
Indvandrede i 2001-2002 .....	15
Indvandrede i 2003.....	15
Indvandrede i 2004 .....	15
Indvandrede i 2005-2006.....	15
Indvandrede i 2005-2016 .....	15
Indvandrede i 2016-2017 .....	16
Indvandrede i 2018 .....	16
Indvandrede i 2016-2019 .....	16
Indvandrede i 2019-2020 .....	16

## Undersøgelse af indvandrernes medbragte uddannelse

Den første spørgeskemaundersøgelse om indvandrernes medbragte uddannelse blev gennemført i 1999. Herefter blev der gennemført regelmæssige undersøgelser med et lignende undersøgelsesdesign frem til 2006. Siden 1999 har der derfor indgået selvrapporterede oplysninger om medbragt uddannelse i statistikken for befolkningens højst fuldført uddannelse.

Undersøgelserne har været gennemført med ekstern finansiering, da uddannelsesstatistikken i øvrigt vedrører uddannelsesaktiviteter gennemført ved danske uddannelsesinstitutioner.

For perioden 2007-16 blev der ikke gennemført løbende undersøgelse af indvandrernes medbragte uddannelse, hvilket medførte, at andelen af indvandrere med uoplyst uddannelsesoplysning i statistikken steg væsentligt i denne periode. Nogle personer, der er indvandret i denne periode, vil dog indgå i statistikken med en oplyst uddannelse, hvis de fx har taget en uddannelse ved en dansk uddannelsesinstitution, eller hvis uddannelsesoplysningerne er indhentet i forbindelse med autorisering eller lignende.

Danmarks Statistik gennemførte i 2016 en undersøgelse af indvandrernes medbragte uddannelse med støtte fra Rockwoolfonden. Undersøgelsen dækkede personer som er indvandret i perioden 2005-2016 og ikke havde svaret på en tidligere undersøgelse.

I 2018 er der gennemført en undersøgelse som dækker personer som er indvandret i perioden 2016-2017.

### Oplysninger om undersøgelser af indvandrernes medbragte uddannelse

	Spurgte	Besvaret	pct.
	antal		
Indvandret før 1999 <sup>1</sup>	152.154	75.545	50
Indvandret i 1999 og 2000	32.399	27.836	86
Indvandret i 2001 og 2002	17.893	6.615	37
Indvandret i 2003	9.011	3.340	37
Indvandret i 2004	9.559	5.304	55
Indvandret i 2005 og 2006	16.929	13.725	81
Indvandret i 2005-2016 <sup>2</sup>	176.142	65.297	37
Indvandret i 2016 og 2017	54.177	12.439	23
<b>i alt</b>	<b>468.264</b>	<b>210.101</b>	<b>45</b>

<sup>1</sup> Offentliggjort i 2000

<sup>2</sup> Offentliggjort i 2017. Øvrige undersøgelser er løbende offentliggjort med statistikken "Befolkningens Højst fuldførte uddannelse". Kilde: Danmarks Statistik.

Enkelte af undersøgelserne handler om personer indvandret i de samme år, men de enkelte personer har kun besvaret én undersøgelse.

I alt har omkring 468.000 personer indgået i undersøgelserne af indvandrernes medbragte uddannelse fra 1999 til 2017. Heraf har omkring 210.000 besvaret spørgeskemaet og deres selvrapporterede oplysninger er som udgangspunkt anvendt til statistikken. For de resterende omkring 258.000 personer, der ikke har besvaret nogle af undersøgelserne, er uddannelsesoplysningerne imputeret.

Imputering (eller imputation) er betegnelsen for den statistiske proces, hvorved manglede observationer erstattes med sandsynlige værdier efter en bestemt algoritme. Disse værdier kaldes imputerede værdier. De metoder, som er anvendt

til imputation er ændret i takt med metodeudviklingen på området hen over undersøgelserne af indvandrernes medbragte uddannelse – se beskrivelse af de enkelte undersøgelser for detaljer.

En del af de selvrapporterede og imputerede oplysninger erstattes efterfølgende, hvis fx en person tager en uddannelse ved en dansk uddannelsesinstitution, eller hvis uddannelsesoplysninger indhentes i forbindelse med autorisering eller lignende. En betydelig del af indvandrerne udvandrer også igen, hvorefter personerne udgår af statistikken.

### **Betydningen af spørgeskemadata for kvaliteten af uddannelsesstatistikken**

Spørgeskemaundersøgelserne af indvandrernes medbragte uddannelse blev gennemført til statistikformål, dvs. til at give et retvisende billede af uddannelsesniveaut for befolkningsgrupper. Samlet set udgør spørgeskemaundersøgelserne af indvandrernes medbragte uddannelse en begrænset del af det samlede datamateriale, som statistikken over højest fuldført uddannelse er baseret på. Derfor har usikkerhed i spørgeskemaundersøgelserne af indvandrernes medbragte uddannelse heller ikke afgørende indflydelse på det samlede billede af uddannelsesniveaut i Danmark. Hvis spørgeskemadata ikke indgik i statistikken, ville det dog medføre at statistikken ville give et mindre fyldestgørende billede af befolkningens uddannelsesniveaut, fordi der ville mangle oplysninger om en del af befolkningen.

Når statistikken opgøres på mindre persongrupper eller på kommuner, vil usikkerheden være større, end når der ses på hele befolkningen. Usikkerheden afhænger af størrelsen af de grupper eller kommuner, som sammenlignes og hvor stor en andel af oplysningerne, der stammer fra undersøgelserne af indvandrernes medbragte uddannelse. Usikkerheden afhænger også af, hvordan uddannelsesoplysningerne anvendes. Hvis statistikken kun anvendes til at opgøre andelen af personer, der har en erhvervskompetencegivende uddannelse, vil usikkerheden være mindre, end hvis man opgør andelen med en specifik uddannelse fx forskeruddannelse. Trods usikkerheden vil oplysningerne fra spørgeskemaundersøgelserne give et mere retvisende billede af uddannelsesniveaut end, hvis man ikke bruger disse oplysninger.

De metoder, som Danmarks Statistik har anvendt til undersøgelserne af indvandrernes medbragte uddannelse, er sammenlignelige med de metoder, som anvendes i andre lande, hvor uddannelsesstatistikken alene er baseret på data fra spørgeskemadata.

Danmarks Statistik udarbejder også andre statistikker, der alene er baseret på spørgeskemadata, fx Arbejdskraftundersøgelsen, som er det danske bidrag til EU's Labour Force Survey, og indgår i EU's og OECD's tal for arbejdsledighed.

## Undersøgelser dækkende personer indvandret i årene 1970-2006

Målet for undersøgelserne var at indsamle oplysninger fra den del af de indvandrede i et givet år, som opfyldte følgende kriterier:

- på indvandringstidspunktet var i alderen 16-59 år
- ikke var i gang med en uddannelse i Danmark
- ikke havde fuldført en uddannelse i Danmark
- havde opholdstilladelse i Danmark, dvs. registreret med CPR-nr.
- stadig boede i landet, når undersøgelsen gennemføres

Oplysningerne blev dels indhentet hos sprogcentre og dels hos Danmarks Statistik i form af webspørgeskema, papirspørgeskema eller telefoninterviews.

### Dataindsamling fra sprogcentre

Integrationsministeriet og Styrelsen for IT og Læring (STIL)<sup>1</sup> stod for dataindsamlingen fra sprogcentre. De varetog i fællesskab samarbejdet med sprogcentre og koordinerede indsamlingen af data fra alle landets sprogcentre, som efterfølgende blev samlet og videresendt til Danmarks Statistik. STIL sørgede for, at der ikke var dobbeltindberetninger på personniveau, samt at data blev modtaget fra alle sprogcentre.

### Dataindsamling hos Danmarks Statistik

Personer, som ikke havde deltaget i undersøgelsen gennem sprogcentre, blev kontaktet af Danmarks Statistik. I de tidlige undersøgelser blev oplysningerne udelukkende indhentet via papirspørgeskema. I undersøgelsen for indvandrede i årene 2005-2006 blev det indført, at man kunne svare enten via webspørgeskema eller gennem telefoninterview. De personer, der ikke opnåede kontakt med gennem disse muligheder, fik tilsendt et papirspørgeskema med posten.

### Imputering – korrektion for manglende besvarelser

For personer, hvor der ikke gennem undersøgelsens datakilder er fundet en oplysning om uddannelsesniveau, er manglende oplysninger om uddannelsesniveau imputeret. Der er foretaget en enkelt imputation for hver person. Imputationsmetoden blev revideret i 2004. Imputationen er foretaget på baggrund af en logistisk regressionsmodel, der på baggrund af den del af populationen af indvandrere, som har afgivet oplysninger om deres uddannelsesniveau via spørgeskemaet. Imputationen er foretaget separat for de 18 til 21-årige og for personer over 22 år. I imputationsmodellen indgår oplysninger om køn, alder, oprindelsesland og arbejdsfunktionskode (DISCO-kode). For personer over 22 år indgår desuden oplysninger om bruttoindkomst og branche.

---

<sup>1</sup> Tidligere UNI•C

## Undersøgelse dækkende personer indvandret i årene 2005-2016

Undersøgelsen om indvandreres medbragte uddannelse bestod af en spørgeskemaundersøgelse, som foregik ude på sprogcentrene samt ved en webundersøgelse gennem e-Boks. Derudover blev der indhentet eksisterende data gennem Styrelsen for Arbejdsmarked og Rekruttering (STAR). Data modtaget fra STAR indeholder oplysninger fra jobcentrene. Her registreres indvandrere og deres CV-oplysninger, herunder uddannelse, når de er ledige og vurderes at være arbejdsmarkedsparete.

Rockwool Fonden har bidraget til finansieringen af dataindsamlingen og efterfølgende analyseret på data. Læs mere om undersøgelsen her:

<https://www.rockwoolfonden.dk/publikationer/ny-stor-dataindsamling-om-indvandreres-medbragte-uddannelse/>

Målet for undersøgelserne var at indsamle oplysninger fra den del af de indvandrede, der har første indvandringsdato den 1. januar 2005 eller senere og opfyldte følgende krav:

- på indvandringstidspunktet var i alderen 15-65 år
- ikke var i gang med en uddannelse i Danmark
- ikke havde fuldført en uddannelse i Danmark
- ikke var del af kvalifikationsregisteret
- ikke havde deltaget i en tidligere undersøgelse om medbragt uddannelse
- havde opholdstilladelse i Danmark, dvs. registreret med CPR-nr.
- stadig boede i landet, når undersøgelsen gennemføres

Til undersøgelsen blev der udarbejdet forskellige skemaer. Der findes i alt 16 landespecifikke skemaer, der tager afsæt i de pågældende landes uddannelsessystemer, ligesom der er udarbejdet et skema på dansk/engelsk. Landeskemaerne er formuleret på de respektive landes sprog og tilpasses landets uddannelsessystem. Skemaerne er udarbejdet på denne måde for at gøre det lettere for især relativt nyankomne indvandrere at besvare skemaet uden at skulle forholde sig til det danske uddannelsessystem, herunder uddannelsesbetegnelser og uddannelsesniveauer

Under udarbejdelsen af skemaerne har Danmarks Statistik været i kontakt med flere institutioner med erfaringer med dataindsamling blandt indvandrere. Endvidere er dataindsamlingen og skemaer drøftet med det svenske og det norske statistikbureau. Begge lande har gennemført lignende undersøgelser.

### Dataindsamling på sprogcentrene

Ved at foretage en spørgeskemaundersøgelse på sprogcentrene var det forventet, at ramme en bredere gruppe af indvandrere, især de mere ressourcetsvage, som ellers forventes at blive underrepræsenteret i statistikken.

Dataindsamlingen på landets sprogcentre blev gennemført af flere omgange. Først blev der gennemført en kort pilottest på enkelte hold. Formålet med dette var at få afprøvet de første skemaer, hvor lang tid det tog at udfylde skemaerne, samt feedback fra kursisterne og undervisere.

Den anden pilottest blev gennemført med det formål at få afprøvet forskellige indsamlingsmuligheder. Følgende muligheder blev afprøvet:

- Intervieweren kommer ud i klassen. Underviser uddeler skema, men interviewer fungerer som supervisor, som indsamler spørgeskemaerne samt besvarer eventuelle spørgsmål.
- Intervieweren observerer holdet og underviseren, når underviseren uddeler, instruerer og indsamler spørgeskemaerne.
- Interviewer er ikke til stede i klassen under afholdelse af undersøgelsen. Underviseren fungerer derfor i en selvstændig rolle og skal uddele, indsamle og instruerer kursisterne. Hvis der opstår problemer eller spørgsmål kan underviser benytte Danmarks Statistiks Hotline.

Inden den anden pilottest gik i gang modtog Danmarks Statistik et dataudtræk som indeholdt oplysninger om kursisterne. På baggrund af dette datasæt blev der præudfyldt skemaer til den enkelte kursist. Pilottesten bestod af 16 forskellige skemaer; et dansk/engelsk skema samt 15 landespecifikke skemaer. Hvis der ikke fandtes et relevant landespecifikt skema til den enkelte kursist modtog de et skema på dansk/engelsk.

Efter pilottesten var gennemført, evaluerede Danmarks Statistik projektet med sprogcentret, herunder undervisere, interviewer og de kommentarer der kom ind på hotlinen.

Dette medførte blandt andet at der blev tilføjet flere sprog til nogle af skemaerne, ekstra landeskema til Tyrkiet, samt længere tid til at besvare spørgeskemaet i den endelige undersøgelse. Desuden blev det besluttet at gennemføre undersøgelsen uden brug af interviewer.

Efter de to pilottest blev undersøgelsen, med få undtagelse gennemført i resten af landet.

De skoler som deltog i projektet, bruger Ludus som administrativt system. Dette gjorde det muligt at få det dataudtræk, som skulle bruges til at præudfylde skemaerne.

Overordnet skete dataindsamlingen på følgende måde:

- Skolerne blev kontaktet på mail af Danmarks Statistik med kort information om undersøgelse. Herefter blev der aftalt et tidspunkt hvor skolen og Danmarks Statistik over telefon kunne aftale mere.
- Herefter sendte skolerne et dataudtræk til Danmarks Statistik, som så præudfyldte skemaerne.
- Skemaerne blev pakket i kuverter, så hver skole modtog en kasse med et antal kuverter som passede til antal hold. Ud over dette modtog skolen en kuvert med ekstra skemaer samt ekstra skemaer på mail, samt returlabel.

Gennem hele dataindsamlingen på sprogcentre har Danmarks Statistik været afhængige af et godt samarbejde med personalet på sprogcentre. Der skulle med kort varsel sættes tid af til forberedelse, administration og selve undersøgelsen. I hele forløbet er Danmarks Statistik blevet mødt med et stort engagement, velvilje fra både ledelse og undervisere. Sprogcentrenes indsats har derfor været helt central for, at Danmarks Statistik har kunnet gennemføre undersøgelsen.

Alle skemaer er sendt retur til Danmarks Statistik og er følgelig blevet tastet. Undervejs blev bortfaldskategorier tastet. Ved bortfaldskategorier angives grunden til at et skema ikke er udfyldt. Dette kan være grundet sygdom, ikke mødt op, vil ikke deltage, kan ikke læse skemaet osv.

### **Dataindsamling hos Danmarks Statistik**

Udover de førnævnte indsamlinger via sprogcentre og STAR, blev der gennemført en spørgeskemaundersøgelse gennem e-Boks.

Denne blev gennemført ad to omgange. Først en pilottest for at teste skemaet og sidenhen er spørgeskemaet blevet sendt til de personer som ikke har svaret på skemaet via sprogcentre.

Der var mulighed for at vælge mellem at få skemaet på dansk eller på engelsk, og ved spørgsmål til undersøgelsen var det muligt at kontakte Danmarks Statistiks hotline.

Det var forventet, at man gennem e-Boks rammer de mere ressourcestærke indvandrere, da det antages, at man skal have et vist niveau inden for dansk sprogbrug samt kendskab til de offentlige systemer, for at benytte e-boks. Desuden var det forventet, at man rammer indvandrere der har været i landet længere tid, end dem der modtag et skema på sprogcentre.

### **Databehandling**

Data fra de forskellige kilder er løbende kommet ind. Fejlsøgningen er derfor sket løbende. Dette har bl.a. omfattet validering af cpr-numre, vurdering af datakilder samt vurdering af besvarelser. Da undersøgelsen delvist baserer sig på spørgeskemaer, har det ikke været muligt, at indhente oplysninger fra den fulde population. For den resterende del af populationen er der foretaget en imputering af data, således at der bliver givet et kvalificeret skøn for den del af populationen, hvor der ikke foreligger konkrete oplysninger.

### **Imputation – korrektion af manglende besvarelser**

For personer, hvor der ikke gennem undersøgelsens datakilder er fundet en oplysning om uddannelsesniveau, er manglende oplysninger om uddannelsesniveau imputeret. Der er foretaget en enkelt imputation for hver person. Imputationen er foretaget ved hjælp af en algoritme, baseret på random forests.<sup>2</sup> Denne imputation er baseret på oplysninger om henvisningskategori til danskursus, danskursusniveau, partners uddannelse, antal børn, antal personer i familien, familietype, køn, alder, fødeår, oprindelsesland og arbejdsfunktionskode.

### **Validering af imputationen i 2016-undersøgelsen**

For at undersøge præcisionen af imputationen, foretog Danmarks Statistik en supplerende dataindsamling, hvor en del af de tidligere adspurgte, som ikke besvarede undersøgelsen sidst, er blevet spurgt igen. Danmarks Statistik indhentede oplysninger for omkring 3.000 personer og disse oplysninger blev sammenholdt med den imputation der oprindeligt blev foretaget. Denne undersøgelse viste, at:

- 83 pct. af personerne med imputerede uddannelsesoplysninger placeres korrekt i forhold til at have eller ikke have en erhvervskompetencegivende uddannelse.
- Fordelingen på uddannelsesniveauer blandt de 3.000 genadspurgte svarer til fordelingen af imputerede værdier.
- I 33 pct. af tilfældene rammer imputeringen præcist på uddannelsesniveauet fx lang videregående uddannelse.

---

<sup>2</sup> For yderligere information, se Stekhoven, D.J. and Bühlmann, P. (2012): "MissForest - nonparametric missing value imputation for mixed-type data". *Bioinformatics*, 28(1): 112-118.



- Når imputeringen ikke rammer helt rigtigt, så rammer den typisk tæt på, fx bliver en person med en forskeruddannelse (ph.d.) imputeret til at have en lang videregående uddannelse (kandidat) snarere end grundskole eller EUD.

Når uddannelsesstatistikken opgøres på mindre persongrupper eller på kommuner, så vil usikkerheden være større, end når der ses på hele befolkningen. Usikkerheden afhænger af størrelsen af de grupper eller kommuner, som sammenlignes og hvor stor en andel af oplysningerne, der er indhentet gennem spørgeskemaundersøgelser.

## **Undersøgelse dækkende personer indvandret i årene 2016-2018**

Undersøgelsen om indvandreres medbragte uddannelse bestod af en webundersøgelse gennem e-boks, samt data fra Udlændige- og Integrationsministeriet og Styrelsen for Arbejdsmarked og Rekruttering (STAR).

Målet for undersøgelserne var at indsamle oplysninger fra den del af de indvandrede der har første indvandringsdato den 1. juli 2016 eller senere og opfyldte følgende krav:

- på indvandringstidspunktet var i alderen 15-65 år
- ikke var i gang med en uddannelse i Danmark
- ikke havde fuldført en uddannelse i Danmark
- ikke var del af kvalifikationsregisteret
- havde opholdstilladelse i Danmark, dvs. registreret med CPR-nr.
- stadig boede i landet, når undersøgelsen gennemføres

### **Dataindsamling fra sprogcentrene**

Udlændige- og Integrationsministeriet stod for dataindsamlingen fra sprogcentrene. Sprogcentrene indberetter kursisternes uddannelsesniveau, når de bliver tilmeldt danskundervisning. Data blev efterfølgende samlet i en fil, som blev videre sendt til Danmarks Statistik.

### **Dataindsamling hos Danmarks Statistik**

Der blev gennemført en spørgeskemaundersøgelse gennem e-Boks. Der var mulighed for at vælge mellem at få skemaet på dansk eller på engelsk, og ved spørgsmål til undersøgelsen var det muligt at kontakte Danmarks Statistiks hotline.

### **Databehandling**

Data fra de forskellige kilder er løbende kommet ind. Fejlsøgningen er derfor sket løbende. Dette har bl.a. omfattet validering af cpr-numre, vurdering af datakilder samt vurdering af besvarelser. Da undersøgelsen delvist baserer sig på spørgeskemaer, har det ikke været muligt, at indhente oplysninger fra den fulde population. For den resterende del af populationen er der foretaget en imputering af data, således at der bliver givet et kvalificeret skøn for den del af populationen, hvor der ikke foreligger konkrete oplysninger.

### **Imputation – korrektion af manglende besvarelser**

For personer, hvor der ikke gennem undersøgelsens datakilder er fundet en oplysning om uddannelsesniveau, er manglende oplysninger om uddannelsesniveau imputeret. Der er foretaget en enkelt imputation for hver person. Imputationen er foretaget ved hjælp af en algoritme, baseret på random forests<sup>2</sup>. Denne imputation er baseret på oplysninger om henvisningskategori til danskursus, danskursusniveau, partners uddannelse, antal børn, antal personer i familien, familietype, køn, alder ved indvandring, år for først indvandring, opholdsgrundlag, oprindelsesland og arbejdsfunktionskode.

## Undersøgelse dækkende personer indvandret i årene 2016-2019

Undersøgelsen om indvandrerens medbragte uddannelse blev i 2019 foretaget for en udvidet population som følge af Danmarks Statistiks revision af kvalifikationsregistret. Undersøgelsen bestod af data fra: a) Udlændinge- og Integrationsministeriet, b) Styrelsen for Arbejdsmarked og Rekruttering (STAR) samt fra c) en spørgeskemaundersøgelse via e-Boks. Målet for undersøgelsen var at indsamle oplysninger om højest fuldført uddannelse blandt indvandrede personer med ukendt uddannelse.

### Revision af kvalifikationsregistret

I 2019 påbegyndte Danmarks Statistik en revision af kvalifikationsregistret, hvilket medførte ændringer af undersøgelsen om indvandrerens medbragte uddannelse. Disse ændringer havde betydning for undersøgelsens population samt imputation af manglende uddannelsesoplysninger. Konkret blev der foretaget følgende ændringer:

- Populationen udvides fra udelukkende at indeholde indvandrere til også at indeholde danskere og efterkommere med en første indvandringsdato i undersøgelsesperioden
- Populationens aldersgrænse udvides fra 15-65 år til 15-69 år
- Imputation af manglende uddannelsesoplysninger foretages nu for det seneste år samt de tre forrige år

### Populationen

Som følge af revisionen beskrevet i ovenstående afsnit, blev populationen i denne undersøgelse udvidet iht. indvandringstype og aldersgrænse. Endvidere havde den udvidede periode for imputation også betydning for populationen.

Da Danmarks Statistik fra 2019 og fremadrettet foretager imputation af manglende uddannelsesoplysninger for personer indvandret i de seneste fire år, blev det i denne undersøgelse vurderet nødvendigt at korrigere for den udvidede population fire år tilbage i tid for at gøre data sammenlignelige. Populationen for denne undersøgelse indeholder således:

- Personer indvandret mellem 1. januar 2019 og 30. september 2019 som var mellem 15 og 69 år på indvandringstidspunktet
- Indvandrere indvandret mellem 1. januar 2016 og 30. september 2019 som var mellem 66 og 69 år på indvandringstidspunktet
- Danskere og efterkommere indvandret mellem 1. januar 2016 og 30. september 2019 som var mellem 15 og 69 år på indvandringstidspunktet

Alle personer i populationen opfyldte derudover følgende krav:

- havde ikke fuldført en uddannelse i Danmark
- var ikke en del af kvalifikationsregistret
- havde ikke deltaget i en tidligere undersøgelse om medbragt uddannelse
- havde opholdstilladelse i Danmark, dvs. registreret med CPR-nr.
- boede i Danmark på det referencetidspunkt, hvorfra undersøgelsens population blev trukket

Den udvidede population er således gældende fra 1. januar 2016 og frem og har betydning for kvalifikationsregistret fra denne dato.

### **Dataindsamling fra sprogcentre**

I undersøgelsen om indvandrernes medbragte uddannelse blev uddannelsesoplysninger fra de danske sprogcentre anvendt. Dataindsamlingen blev foretaget af Udlændinge- og Integrationsministeriet, som samlede de danske sprogcentres indberetninger af kursisters uddannelsesniveau til en samlet fil og videre sendte denne til Danmarks Statistik.

### **Dataindsamling fra jobcentre**

Ud over data fra sprogcentre blev der også i denne undersøgelse anvendt data fra de danske jobcentre. STAR foretog indsamling af CV-data fra jobcentre indeholdende uddannelsesniveau, samlede disse data til et samlet datasæt og videre sendte dette til Danmarks Statistik.

### **Dataindsamling hos Danmarks Statistik**

For den del af undersøgelsens population, hvortil der ikke kunne knyttes en uddannelsesoplysning fra Danmarks Statistiks højst fuldførte uddannelse-register eller fra de to kilder beskrevet i ovenstående, blev der gennemført en spørgeskemaundersøgelse. Spørgeskemaundersøgelsen blev foretaget af Danmarks Statistik og blev gennemført som et webspørgeskema udsendt via e-Boks. Der var mulighed for at vælge mellem et skema på dansk eller engelsk. Tilmed var der mulighed for at kontakte Danmarks Statistiks hotline ved eventuelle spørgsmål til skemaet.

### **Databehandling**

Data fra de forskellige kilder er blevet modtaget løbende i Danmarks Statistik. Fejlsøgningen af disse data er derfor sket løbende. Fejlsøgningen har bl.a. omfattet validering af cpr-numre, vurdering af datakilder samt vurdering af besvarelser. Da undersøgelsen delvist er baseret på spørgeskemaer, har det ikke været muligt at indhente oplysninger fra den fulde population. For den del af populationen, hvortil det ikke har været muligt at knytte en uddannelsesoplysning fra de anvendte kilder, er der foretaget imputation. Denne imputation af uddannelsesoplysninger bidrager til at give et kvalificeret skøn af uddannelsesniveaet for denne del af populationen.

### **Imputation – korrektion af manglende besvarelser**

For personer, hvor der ikke gennem undersøgelsens datakilder er fundet en oplysning om uddannelsesniveau, er manglende oplysninger om uddannelsesniveau imputeret. Der er foretaget en enkelt imputation for hver person. Imputationen er foretaget ved hjælp af en algoritme, baseret på random forests<sup>2</sup>. Denne imputation er baseret på oplysninger om henvisningskategori til danskursus (i de danske sprogcentre), danskursusniveau, partners uddannelse, antal børn, antal personer i familien, familietype, køn, alder ved første indvandring, år for første indvandring, opholdsgrundlag, oprindelsesland og arbejdsfunktionskode.

Fra 2019 og frem foretages årligt en revidering af imputerede uddannelsesoplysninger for personer indvandret de seneste fire år. Dette skyldes, at datagrundlaget for imputationen styrkes hvert år.

## Undersøgelse dækkende personer indvandret i 2019-2020

Undersøgelsen om indvandrernes medbragte uddannelse blev i 2020 foretaget for personer indvandret i 4. kvartal 2019 eller i 1., 2. eller 3. kvartal 2020. Undersøgelsen tog udgangspunkt i revisionen foretaget i 2019 og de ændringer som medførte heraf. Undersøgelsen bestod af data fra: a) Udlændinge- og Integrationsministeriet, b) Styrelsen for Arbejdsmarked og Rekruttering (STAR) samt fra c) en spørgeskemaundersøgelse via e-Boks. Målet for undersøgelsen var at indsamle oplysninger om højest fuldført uddannelse blandt indvandrede personer med ukendt uddannelse.

### Populationen

Populationen for undersøgelsen indeholder personer indvandret mellem 1. oktober 2019 og 30. september 2020, som var mellem 15 og 69 år på indvandringstidspunktet. Populationen består af både indvandrere, efterkommer og personer med dansk oprindelse.

Alle personer i populationen opfyldte derudover følgende krav:

- havde ikke fuldført en uddannelse i Danmark
- var ikke en del af kvalifikationsregistret
- havde ikke deltaget i en tidligere undersøgelse om medbragt uddannelse
- havde opholdstilladelse i Danmark, dvs. registreret med CPR-nr.
- Havde bopæl i Danmark pr. 30. september 2020

### Dataindsamling fra sprogcentre

I undersøgelsen om indvandrernes medbragte uddannelse blev uddannelsesoplysninger fra de danske sprogcentre anvendt. Dataindsamlingen blev foretaget af Udlændinge- og Integrationsministeriet, som samlede de danske sprogcentres indberetninger af kursisters uddannelsesniveau til en samlet fil og videregav denne til Danmarks Statistik. Grundet forsinket dataleverance af oplysninger vedr. 3. kvartal 2020, blev der i nærværende undersøgelse anvendt data fra de danske sprogcentre dækkende perioden til og med 2. kvartal 2020.

### Dataindsamling fra jobcentre

Ud over data fra sprogcentre blev der også i denne undersøgelse anvendt data fra de danske jobcentre. STAR foretog indsamling af CV-data fra jobcentre indeholdende uddannelsesniveau, samlede disse data til et samlet datasæt og videregav dette til Danmarks Statistik.

### Dataindsamling hos Danmarks Statistik

For den del af undersøgelsens population, hvortil der ikke kunne knyttes en uddannelsesoplysning fra Danmarks Statistiks højest fuldførte uddannelse-register eller fra de to kilder beskrevet i ovenstående, blev der gennemført en spørgeskemaundersøgelse. Spørgeskemaundersøgelsen blev foretaget af Danmarks Statistik og blev gennemført som et webspørgeskema udsendt via e-Boks. Der var mulighed for at vælge mellem et skema på dansk eller engelsk. Tilmed var der mulighed for at kontakte Danmarks Statistiks hotline ved eventuelle spørgsmål til skemaet.

### Databehandling

Data fra de forskellige kilder er blevet modtaget løbende i Danmarks Statistik. Fejlsøgningen af disse data er derfor sket løbende. Fejlsøgningen har bl.a. omfattet validering af cpr-numre, vurdering af datakilder samt vurdering af besvarelser. Da undersøgelsen delvist er baseret på spørgeskemaer, har det ikke været

muligt at indhente oplysninger fra den fulde population. For den del af populationen, hvortil det ikke har været muligt at knytte en uddannelsesoplysning fra de anvendte kilder, er der foretaget imputation. Denne imputation af uddannelsesoplysninger bidrager til at give et kvalificeret skøn af uddannelsesniveaet for denne del af populationen.

#### **Imputation – korrektion af manglende besvarelser**

For personer, hvor der ikke gennem undersøgelsens datakilder er fundet en oplysning om uddannelsesniveau, er manglende oplysninger om uddannelsesniveau imputeret. Der er foretaget en enkelt imputation for hver person. Imputationen er foretaget ved hjælp af en algoritme, baseret på random forests<sup>2</sup>. Denne imputation er baseret på oplysninger om henvisningskategori til danskkursus (i de danske sprogcentre), danskkursusniveau, partners uddannelse, antal børn, antal personer i familien, familietype, køn, alder ved første indvandring, år for første indvandring, opholdsgrundlag, oprindelsesland og arbejdsfunktionskode.

Der blev ifm. denne undersøgelse foretaget en revidering af imputerede uddannelsesoplysninger for personer indvandret de seneste fire år. Denne revidering blev foretaget pba. et stærkere datagrundlag for imputationen med afsæt i nye data.

## **Oversigt over undersøgelserne**

### **Indvandrede efter 1970 og før 1. januar 1999**

Den første undersøgelse blev igangsat i juni 1999 og blev afsluttet medio 2000. Målgruppen var personer indvandret efter 1970, som stadig opholdt sig i Danmark 1. januar 1999. I alt blev 152.154 personer spurgt. 75.545 svarede på undersøgelsen. Ved imputeringen i 2008 blev der imputeret uddannelsesniveau for 72.015 personer.

### **Indvandrede i 1999-2000**

Undersøgelsen blev igangsat i 2002 og afsluttet i 2003. Populationen udgjorde 32.399 personer. Der kom data fra sprogcentre som indeholdt oplysninger om 25.896 personer. Restgruppen på 6.503 personer modtog spørgeskema, som blev besvaret af 1940 personer. Ved imputeringen i 2008 blev der imputeret uddannelsesniveau for 4.550 personer.

### **Indvandrede i 2001-2002**

Undersøgelsen blev igangsat i 2003 og afsluttet i 2004. Pga. problemer med indberetningssystemet, som sprogcentre anvendte ved indberetning til Integrationsministeriet, modtog Danmarks Statistik ikke oplysninger for indvandrede i 2002 og for en del af de indvandrede i 2001 fra sprogcentre. Populationen udgjorde 17.893. Der kom besvarelser fra 6.615 personer. Ved imputering i 2008 blev der imputeret uddannelsesniveau for 11.716 personer.

### **Indvandrede i 2003**

Undersøgelsen blev igangsat i 2005 og afsluttet i 2006. Populationen udgjorde 9.011. Der kom besvarelser fra 3.340 personer. Ved imputering i 2008 blev der imputeret uddannelsesniveau for 5769 personer. I denne undersøgelse indgik der ikke oplysninger fra sprogcentre.

### **Indvandrede i 2004**

Undersøgelsen blev igangsat i 2006 og afsluttet i 2008. Populationen udgjorde 9.559 personer. Der kom data fra sprogcentre som indeholdt oplysninger om 3.681 personer. Restgruppen på 5.878 personer modtog spørgeskema, som blev besvaret af 1.623 personer. Ved imputering i 2008 blev der imputeret uddannelsesniveau for 4.465 personer.

### **Indvandrede i 2005-2006**

Undersøgelsen blev igangsat i 2009 og afsluttet i 2010. For første gang blev der oprettet mulighed for besvarelse på nettet. Populationen udgjorde 16.929 personer. Der kom data fra sprogcentre som indeholdt oplysninger om 9.074 personer. Restgruppen på 7.855 personer modtog spørgeskema, som blev besvaret af 4.651 personer, hvilket medførte at der blev imputeret værdier for 3.204 personer.

### **Indvandrede i 2005-2016**

Undersøgelsen blev igangsat i 2016 og afsluttet i 2017. Indsamlingen foregik via sprogcentre, digital post samt eksisterende data fra Styrelsen fra Arbejdsmarked og Rekruttering. Populationen udgjorde 176.142 personer. For 65.297 er der indsamlet uddannelsesoplysninger fra forskellige kilder, hvilket medførte at der blev imputeret uddannelsesniveau for 110.845 personer.

### **Indvandrede i 2016-2017**

Undersøgelsen blev igangsat i 2017 og afsluttet i 2018. Danmarks Statistik modtog data fra Udlændige- og Integrationsministeriet, digital post samt eksisterende data fra Styrelsen for Arbejdsmarked og Rekruttering. Populationen udgjorde 54.177 personer. For 12.439 er der indsamlet uddannelsesoplysninger fra forskellige kilder, hvilket medførte at der blev imputeret uddannelsesniveau for 41.738 personer.

### **Indvandrede i 2018**

Undersøgelsen blev igangsat i 2018 og afsluttet i 2019. Danmarks Statistik modtog data fra Udlændige- og Integrationsministeriet, digital post samt eksisterende data fra Styrelsen for Arbejdsmarked og Rekruttering. Populationen udgjorde 39.981 personer. For 11.167 er der indsamlet uddannelsesoplysninger fra forskellige kilder hvilket medførte at der blev imputeret uddannelsesniveau for 28.814 personer.

### **Indvandrede i 2016-2019**

Undersøgelsen blev igangsat i 2019 og afsluttet i 2020. Danmarks Statistik modtog data fra Udlændige- og Integrationsministeriet, Styrelsen for Arbejdsmarked og Rekruttering samt fra en spørgeskemaundersøgelse via e-Boks. Populationen udgjorde 40.497 personer. For 9.439 personer blev der indsamlet uddannelsesoplysninger fra de førnævnte kilder. For de resterende 31.058 personer blev uddannelsesniveaet imputeret.

I 2019 foretog Danmarks Statistik en revision af kvalifikationsregistret, hvilket havde betydning for populationen undersøgelsen om indvandrernes medbragte uddannelse. I 2019 blev populationen dermed udvidet til at indeholde alle indvandrede personer i alderen 15-69 år. Denne udvidelse blev foretaget med tilbagevirkende kraft mhp. at korrigere de seneste fire år således at populationerne blev sammenlignelige.

### **Indvandrede i 2019-2020**

Undersøgelsen blev igangsat i 2020 og afsluttet i 2021. Danmarks Statistik modtog data fra Udlændige- og Integrationsministeriet, Styrelsen for Arbejdsmarked og Rekruttering samt fra en spørgeskemaundersøgelse via e-Boks. Populationen udgjorde 35.022 personer. For 11.601 personer blev der indsamlet uddannelsesoplysninger fra de førnævnte kilder. For de resterende 23.421 personer blev uddannelsesniveaet imputeret.