

## Opholdstilladelser på individniveau fra 1997

### Baggrund

Danmarks Statistik har fra Udlændingestyrelsens sagsregister modtaget udlændingesager om førstegangstilladelser til ophold i Danmark med cpr-nummer fra 1997 og frem, og modtager løbende henholdsvis foreløbige og endelige opholdsdata kvartalsvist og årligt. Personer der indvandrer til Danmark, som ikke har nordisk eller dansk statsborgerskab, skal have en opholdstilladelse for at kunne lade sig indregistrere i folkeregistret.

Ved registrering i folkeregistret, bliver personens udlændingesagsnummer koblet til CPR-nummeret, og det er derfor muligt at danne et statistikregister med indvandrede opholdstilladelser. Det har dog vist sig, at ikke alle indvandrede personer med ikke-nordisk/dansk statsborgerskab fra 1997 og frem, har fået koblet et cpr-nummer til personens sag i Udlændingestyrelsens sagsregister. Andelen af personer der årligt indvandrer, som ikke har fået koblet en opholdstilladelse svinger mellem 6 og 28 pct. i perioden 1997-2020, samlet set gælder det 16 pct. af de indvandrede i perioden (se tabellerne sidst i dette dokument). Danmarks Statistik har derfor valgt at imputere data, for de personer, der er indvandret fra 1997 og frem, som burde have fået koblet en opholdstilladelse på. For personer der har seneste indvandring før 1997, er data om opholdstilladelse dannet på anden måde, som er beskrevet i på [emnesiden vedrørende indvandrere og efterkommere](#).

Danmarks Statistik har udgivet disse data siden 2014, men har nu foretaget en revidering af metoden til at danne data, og det er frem over disse data, der vil blive anvendt til opgørelser over indvandreres opholdstilladelse.

### Opholdstilladelser tildelt fra Udlændingestyrelsen sammenholdt med indvandrede opholdstilladelser

Det årlige antal offentliggjorte førstegangstilladelser fra Udlændingestyrelsen, kan ikke direkte sammenlignes med antal og fordeling af indvandrede personers opholdstilladelser.

Begrebet ”førstegangstilladelse” er i Udlændingestyrelsen knyttet til ”sagsområder” dvs. kategorier af tilladelser og ikke til personer, der således kan have flere ”førstegangstilladelser”. Ved en ”førstegangstilladelse” forstås en tilladelse, der gives for første gang til en person inden for en af de forskellige kategorier af tilladelser (sagsområder). Fornyelser af samme type opholdstilladelse opgøres derfor ikke som ”førstegangstilladelse”, mens en opholdstilladelse inden for et andet sagsområde (af en anden kategori) vil blive klassificeret som en ”førstegangstilladelse” selv om personen tidligere har haft en anden type af tilladelse.” Den samme person kan derfor have flere ”førstegangstilladelser”.

Udlændingestyrelsen opgør hvert år det samlede antal førstegangstilladelser, der er tildelt i løbet af året. Førstegangstilladelser kan tildeles personer:

- der ikke er indvandret til Danmark, men som måske eller måske ikke efterfølgende indvandrer til Danmark – i disse tilfælde vil der derfor ikke indvandre et lige så højt antal personer, som antallet af personer med bopæl i udlandet, der har fået tildelt en opholdstilladelse;
- personer i asylsystemet, som endnu ikke er registreret i folkeregistret – disse personer vil oftest blive indregistreret i folkeregistret umiddelbart efter at have modtaget tilladelsen, og medfører derfor en efterfølgende indvandring – samt;
- personer der allerede er indregistreret i folkeregistret, og som har en anden type opholdstilladelse. Personer der har opholdstilladelse, og har bopæl i Danmark, vil med mellemrum få fornyet deres opholdstilladelse. Hvis en person får fornyet en familiesammenføringstilladelse, vil fornyelsen ikke fremgå af opgørelsen. Hvis en person skifter fra at have familiesammenføringstilladelse til fx opholdstilladelse til asyl mv., vil Udlændingestyrelsen offentliggøre den nye opholdstilladelse som en førstegangstilladelse, og Danmarks Statistik modtager samtidig oplysningen om en ny førstegangstilladelse. Førstegangstilladelser til personer med bopæl i Danmark, som modtager en ny type førstegangstilladelse, vil derfor ikke medføre en indvandring.
- Dertil kan der være tidsforskydning mellem tildeling af tilladelse og indvandring til Danmark.

Der er dermed flere grunde til, at antallet af tildelte opholdstilladelser i et år, ikke direkte kan sammenholdes med antallet af indvandrede som skal have en opholdstilladelse.

## **Dokumentation**

Denne dokumentation beskriver, hvordan data om opholdsgrundlag for indvandrede, der er indvandret fra 1997 og frem er dannet ud fra oplysninger fra Udlændingestyrelsen, kombineret med imputering ved hjælp af CPR-oplysninger, arbejdsmarkeds- og uddannelsesoplysninger. Endvidere beskrives de overordnede metodiske valg, som Danmarks Statistik har truffet i denne forbindelse samt begrundelsen herfor. Endelig beskrives, hvordan data anvendes til at danne register over indvandreres første og seneste opholdstilladelse, på et givent statustidspunkt, samt hvilke forskelle der er i de reviderede data i forhold til de tidligere udgivelser.

## **1. Overordnede metodiske valg**

### **1.1. Beskrivelse af alle individer i populationen**

Det har som udgangspunkt været en beslutning, at arbejdet med at danne et register med opholdstilladelse for personer der er indvandret fra 1997 og frem, skulle munde ud i et register, der angiver en opholdstilladelse for samtlige individer i populationen. Det skyldes behovet for at kunne koble oplysningerne for samtlige individer i indvandrerpopulationen med andre personstatistiske registre i forbindelse med statistiske analyser. En stor andel uoplyste, begrænser muligheden for at anvende opholdsgrundlag i den statistiske belysning af indvandrerpopulationen.

Til beskrivelse af opholdstilladelse for samtlige i populationen, er der anvendt en kombination af opholdstilladelser fra Udlændingestyrelsen og imputerede opholdstilladelser. Andelen af imputerede i perioden 1997-2019 er samlet set 16 pct., hvilket dækker over årlige imputeringer, der svinger mellem 6 og 28 pct.

Ulempen ved at anvende imputering er, at data ikke er 100 pct. præcis på individniveau og at analyseresultaterne derfor skal tages med forbehold, særligt når der ses på mindre populationer. Fordelen er imidlertid, at det på denne måde bliver muligt at anvende data til analyser, hvori der indgår data om forskellige forhold som fx uddannelse, arbejdsmarkedstilknytning m.v. Men det er vigtigt at være opmærksom på, at data ikke uden forbehold kan anvendes til analyser af fx mindre geografiske områder, og at den statistiske usikkerhed især skal medtages i konklusioner ud fra data om opholdsgrundlag, når der opdeles på mindre grupper.

## 1.2 Opholdstilladelse ved indvandring

Data om opholdsgrundlag for indvandrere med en indvandring fra 1997 og frem, bliver som udgangspunkt dannet i forbindelse med indvandringstidspunktet. Disse data er lagt i registret d101200.psd\_indvandrede\_opholdsgrundlag. Det er dannelsen af opholdstilladelse ved indvandring, som bliver beskrevet i denne dokumentation. Der bliver derudover dannet et afledt register, der opgør den første og den seneste opholdstilladelse pr. 31.12 fra 2004 og frem (d101200.psd\_status\_opholdsgrundlag). Der er en kort beskrivelse af dannelsen af dette register i slutningen af dette dokument (afsnit 6)

## 1.3. Behandling af opholdstilladelser fra Udlændingestyrelsen

I udarbejdelsen af registret med opholdstilladelse for indvandrede fra 1997, er data fra Udlændingestyrelsen anvendt efter bestemte afgrænsninger.

Danmarks Statistik modtager oplysning om opholdstilladelser med CPR-nummer på tre niveauer af opholdsgrundlaget, nemlig "Kategori", "Grundlag" og "Forklar", udlændingenummer, samt en tilladelsesdato. Når DST modtager data, bliver CPR-nummeret erstattet med et afidentificeret person\_id, og data bliver lagt i et baggrundsregister med samtlige opholdstilladelser fra Udlændingestyrelsen. Opholdstilladelser uden CPR-nummer bliver ligeledes lagt i registret.

Kun tilladelser med tilladelsesdato indenfor 15 måneder før og 30 dage efter indvandring bliver koblet på personer der indvandrer. Hvis der er mere end én tilladelse i denne periode, vælges den med færrest dage mellem indvandring og tilladelsesdato. 95 pct. af de anvendte tilladelser har en tilladelsesdato der ligger inden for tre måneder før og 30 dage efter indvandring. Hvis der er mere end én tilladelse pr. dag pr. person, bliver opholdstilladelsen sorteret stigende efter "Kategori", "Grundlag", "Forklar". Afgrænsningen til tilladelser, der ligger mindre end 15 måneder før indvandringsdatoen, er foretaget ud fra en betragtning om, at undgå at medtage tilladelser fra tidligere indvandring, som ikke længere gælder, og samtidig en vurdering af, at de imputerede tilladelser fra det seneste indvandringstidspunkt, er mere retvisende end en for gammel opholdstilladelse.

Det overordnede resultat er tjekket i forhold til anden viden om omfanget af hvilke opholdstilladelser der er tildelt til udenlandske statsborgere, se b.l.a. [www.statistikbanken.dk/VAN66](http://www.statistikbanken.dk/VAN66). Det er som tidligere nævnt ikke muligt at sammenligne antallet af tildelte tilladelser i ét kalenderår med antallet af indvandrede efter opholdstilladelse, da ikke alle tilladelser medfører en efterfølgende indvandring, og nogle tilladelser tildeles personer, der allerede har bopæl i Danmark, men med en anden type opholdstilladelse. Krydstjekket af indvandrede opholdstilladelse i forhold til antallet af tildelte tilladelser, viser at fordelingen er konsistent over årene, og bekræfter samtidig et ræsonnement om, at der er færre der indrejser ved tildeling af tilladelser til mere midlertidige indvandringsårsager som arbejde- og studie, end ved mere permanente indvandringsårsager som asyl og familiesammenføring.

## 2. Kilder og variable til imputering

Til at danne registret er der anvendt forskellige kilder som imputeringsvariable, til imputering af data om opholdstilladelse for de personer i populationen, der ikke har fået koblet en opholdstilladelse fra Udlændingestyrelsen. De anvendte kilder og deraf afledte imputeringsvariable beskrives i det følgende. Kilderne til variablene er CPR-oplysninger, opholdstilladelser fra Udlændingestyrelsen, Elevregistret, Registerbaseret arbejdsmarkedsstatistik og Lønmodtagerbeskæftigelsen. De afledte variable der anvendes til imputeringen er første indvandringsdato, statsborgerskab ved første indvandring, familierelation ved indvandring, køn, indvandringsalder, nuværende alder, opholdstilladelse fra Udlændingestyrelsen, to cifret uddannelseskode ved første kontakt med Elevregistret, arbejdsmarkedsstatus første og andet år efter indvandring samt lønmodtagerdata om antal lønmodtagertimer og timeløn ved de job, der påbegyndes inden for tre måneder efter indvandringen.

### 2.1. CPR-oplysninger

*CPR-oplysninger – indvandringsdato* Indvandringsdato er Haendelsesdato i registret over indvandring i Danmarks Statistiks personstatistikregister (PSD), d101200.psd\_indvandring.

*CPR-oplysninger – statsborgerskab ved indvandring* Da det er statsborgerskabet ved indvandring, der i høj grad er definerende for opholdstilladelsestypen, har vi anvendt det statsborgerskab, der findes i CPR-registret for personen ved indvandringen, som også hentes fra d101200.psd\_indvandring. Dvs. at selv om personen sidenhen har erhvervet dansk statsborgerskab, er det ikke det, der anvendes som imputeringsvariabel. Hvis der ikke er et oplyst statsborgerskab ved indvandring, anvendes det første statsborgerskab, der er registreret efter personens indvandring eller fødeland.

*CPR-oplysninger – familierelation* Ud fra forskellige oplysninger om egne og familiemedlemmers indvandringstidspunkt, statsborgerskab og evt. oplysninger om opholdsgrundlag, er der dannet en oplysning, om indvandreres evt. familierelation i Danmark ved indvandringen.

I den følgende beskrivelse af dannelse af familierelation kaldes personen, der indvandrer, for hovedpersonen, og det familiemedlem der dannes relation til bliver kaldt for familiemedlemmet, ægtefællen eller forælderen. Bemærk at familierelationer der er indvandret til Danmark efter hovedpersonens egen indvandring ikke bliver medtaget her, da de ikke vil have haft en indflydelse på, om hovedpersonen evt. er blevet familiesammenført.

Først er der dannet en oplysning om, hvorvidt der har været en ægtefælle i Danmark ved indvandringen. Der er her foretaget den skelnen, at det kun er personer der var 18 år ved indvandringen, som kunne have en ægtefælle i Danmark ved indvandringen.

Herefter er der dannet en oplysning om, hvorvidt der har været en forælder ved indvandringen, eller som er indvandret samtidig med personer, der var under 18 år ved indvandringen.

Det har endvidere været forsøgt at koble oplysninger om evt. børn i Danmark ved indvandringen, for personer der var 60 år og derover ved indvandringen. Men der var meget få af disse indvandrere, der fik koblet data på, og det er derfor besluttet, ikke at anvende CPR-data om familierelationer til at definere, om personer på 60 år og derover, er blevet familiesammenført til et barn.

I variabelen Familierelation, bliver familiemedlemmets status som referenceperson defineret ud fra den status familiemedlemmet har, på det tidspunkt hovedpersonen indvandrer. Dvs. hvis familiemedlemmet er dansk statsborger på hovedpersonens indvandringstidspunkt, er referenceperson dansk/nordisk statsborger, uanset om familiemedlemmet var flygtning ved egen indvandring. Familiemedlemmet defineres som enten ”dansk/nordisk statsborger”, ”flygtning” eller ”udlænding, men ikke flygtning”, som svarer til Udlændingestyrelsens nuværende fordeling på referencepersoner.

#### Værdier i variabelen Familierelation:

Kode	Tekst
1	Ægtefælle ved indvandring dansk/nordisk statsborger
2	Ægtefælle ved indvandring flygtning
3	Ægtefælle ved indvandring udlænding men ikke flygtning
11	Forælder før barn, dansk/nordisk statsborger,
12	Forælder før barn, flygtning,
13	Forælder, før barn, udenlandsk statsborger men ikke flygtning
14	Forælder sammen med barn, dansk/nordisk statsborger,
15	Forælder sammen med barn, flygtning,
16	Forælder, sammen med barn, udenlandsk statsborger men ikke flygtning
9	Ingen familierelationer ved indvandring

Udover at blive anvendt som imputeringsvariabel, bliver familierelationen brugt til at danne referencepersonens status som enten ”flygtning”, ”dansk/nordisk statsborger” eller ”udlænding, men ikke flygtning”, for familiesammenførte der er indvandret før 2010. I tilladelser fra før 2010, er der ikke angivet status for referenceperson i ”Forklar”, som i tilladelser tildelt fra 2010. Ved imputering af familierelation er det muligt at sammenholde tilladelser fra før og efter 2010, og fx opgøre hvor mange personer, der er familiesammenført til flygtninge.

CPR-oplysninger –  
øvrige

Køn, indvandringalder, beregnet ud fra første indvandring, samt nuværende alder, bliver anvendt som imputeringsvariable.

## 2.2. Data fra Udlændingestyrelsen

Data om opholdstilladelse fra Udlændingestyrelsen er grundlaget til at kunne danne registeret med opholdstilladelser for indvandrede siden 1997. De indvandrede, der ikke har fået koblet en opholdstilladelse fra Udlændingestyrelsen, får en imputeret opholdstilladelse, med samme variable og værdier som i Udlændingestyrelsens opholdstilladelser.

Opholdstilladelse fra Udlændingestyrelsen er fordelt på tre niveauer, ”Kategori”, ”Grundlag” og ”Forklar”, der gør det muligt at gruppere dem på et meget overordnet niveau, som fx flygtning, familiesammenført, EU/EØS-opholdsbeviser, arbejde, studier, og på et mere detaljeret niveau, som fx familiesammenført til ægtefælle der er flygtning, eller EU/EØS tilladelse til uddannelse. Dertil anvendes tilladelsesdatoen. Tilladelsesdatoen sættes til 01.01.1600, hvis opholdstilladelsen bliver imputeret.

## 2.3. Kontakt med uddannelsessystemet

Fra Elevregistret er personens første kontakt med uddannelsessystemet i indvandringsåret. Den tocifrede uddannelseskode samt dato for start, er blevet anvendt som imputeringsvariabel.

## 2.4. Arbejdsmarkedsstatus

Fra RAS (Registerbaseret Arbejdsmarkedsstatistik) er socioøkonomisk status ultimo november i indvandringsåret og året efter indvandringen anvendt som imputeringsvariable. Socioøkonomisk status er opgjort som beskæftiget, arbejdsløs eller uden for arbejdsmarkedet. Det er kun for indvandringsårene til og med 2018, at begge variable kan dannes på opgørelsestidspunktet. For indvandringsåret 2019 er det kun opgørelsen fra indvandringsåret der kan anvendes, og fra 2020 og frem, vil det ikke være muligt at medtage oplysninger fra RAS, da registeret ikke er dannet, når imputeringen af data dannes tre måneder efter indvandringsåret.

## 2.5. Lønmodtager data

Data om lønmodtagerjob med startdato inden for tre måneder efter indvandringen anvendes. Det er antallet af lønmodtagertimer pr. uge, samt timeløn beregnet i forhold til betalte timer i perioden, der anvendes som imputeringsvariable. Lønmodtagerdata anvendes fra indvandringsåret 2008 og frem.

# 3. Imputering

## 3.1. Metode

Der er foretaget en imputering ud fra metoden ”missForest”, med udgangspunkt i de tilgængelige imputeringsvariable. Personer der er indvandret i et år, som ikke har fået koblet en opholdstilladelse fra Udlændingestyrelsen, har fået imputeret en opholdstilladelse.

MissForest imputering er en ikke-parametrisk variant af Random Forest der kan bruges på de mixed data. Algoritmen bag metoden bearbejder de tilgængelige data for alle individer i datasættet, både dem med og uden opholdsgrundlag, og imputerer herefter ukendte oplysninger (her opholdsgrundlag) på baggrund af kendte (imputeringsvariable). MissForest algoritmen er dokumenteret i Stekhoven Buehlmann (2012)<sup>1</sup> og teknisk implementeret i R (Stekhoven, 2013<sup>2</sup>).

De anvendte imputeringsvariable er som tidligere nævnt første indvandringsdato, statsborgerskab ved første indvandring, familierelation ved indvandring, køn, indvandringsalder, nuværende alder, opholdstilladelse fra Udlændingestyrelsen, to-cifret uddannelseskode ved første kontakt med Elevregistret i indvandringsåret, arbejdsmarkedsstatus første og andet år efter indvandring frem til 2018, samt lønmodtagerdata om antal lønmodtagertimer og timeløn ved de job, der påbegyndes indenfor tre måneder efter indvandringen.

## 3.2. Imputeringspopulation

Imputeringspopulationen består af de af indvandrere, som er indvandret siden 1997, som ikke har et dansk/nordisk statsborgerskab ved indvandring, og som ikke får koblet en opholdstilladelse fra Udlændingestyrelsen ved hjælp af CPR-nummer. Andelen af personer i gruppen af indvandrede, der ikke har et dansk/nordisk statsborgerskab, som ikke får koblet et opholdsgrundlag, svinger mellem 6 og 28 pct. i perioden 1997-2019. Samlet set er det 16 pct. der får en imputeret opholdstilladelse i periode.

---

<sup>1</sup> Stekhoven, D.J. and Buehlmann, P. (2012), 'MissForest – nonparametric missing value imputation for mixed-type data', *Bioinformatics*, 28(1) 2012, 112-118, doi: 10.1093/bioinformatics/btr597

<sup>2</sup> Daniel J. Stekhoven (2013). missForest: Nonparametric Missing Value Imputation using Random Forest. R package version 1.4. <https://CRAN.R-project.org/package=missForest>

Efter imputering af opholdstilladelse, er der foretaget en ekstra imputering af familierelation, for de personer, der er indvandret før 2010, som har en opholdstilladelse til familiesammenføring, hvis det ikke har været muligt at koble familierelation på fra CPR.

### 3.3. Imputeringssikkerhed

Ved missForest-imputeringen af basispopulationen er der en OOBERROR (Out of Bag Error) mellem 18 og 28 pct. i perioden 1997-2019.

Dvs. at programmet vurderer, at modellen rammer rigtigt i mellem 72 pct. og 88 pct. af de imputerede opholdstilladelser. Hvis antallet af imputerede der skønnes ikke at ramme rigtigt, beregnes i forhold til hele populationen, kan det konstateres at mellem 93 og 99 pct. af opholdstilladelserne i registret er korrekte på individniveau i 1997-2019. Gennemsnittet for hele perioden er 96 pct. De sidste 4 pct. bør også betragtes som relativt sikre på et overordnet niveau, fx landsniveau, men der kan typisk være byttet rundt på opholdstilladelser blandt personer med samme karakteristika.

Imputeringen er foretaget med henblik på at få et retvisende billede af indvandrernes opholdsgrundlag på landsplan. Af denne grund har målet alene være at få en lav Out of Bag Error samlet set for landet og der ikke taget højde for at imputeringen samtidigt skal give retvisende billede på eksempelvis på mindre geografiske områder. Dette betyder ikke, at det ikke kan være tilfældet, men jo mindre en population af indvandrere med imputeret opholdsgrundlag der analyseres, desto højere usikkerhed vil der være på fordelingen af deres opholdsgrundlag. Det brugte trænings- og test data setup i denne imputering er ikke tilstrækkeligt til at analysere denne usikkerhed på meget små populationer. Derfor anbefales det at være meget varsom med at udarbejde analyser og konklusioner baseret på små mængder af imputerede opholdsgrundlag.

### 3.4. Efterbehandling

Imputerede opholdstilladelser bliver anvendt direkte som de er. Der er nogle enkelte tilfælde, hvor der ikke er en logisk sammenhæng med den imputerede opholdstilladelse fra bestemte statsborgerskabslande og de opholdstyper der findes fra Udlændingestyrelsen eller surveysvar fra samme land. Det omhandler bl.a. flygtninge og familiesammenførte til flygtninge fra lande der er meget få statsborgere fra. Det skyldes at imputeringen vil søge at finde det bedste resultat for en bestemt kombination af karakteristika, og med meget få personer fra samme statsborgerskabsland, vil statsborgerskabet i sig selv ikke vægte så højt i imputeringen. Da det kun drejer sig om enkelte tilladelser, dvs. omkring ti i alt, og det ikke er muligt at give et bedre bud på opholdstilladelse i disse tilfælde, er det besluttet at beholde disse imputerede tilladelser, uden forsøg på at fejlrette.

## 4. Reviderede data

Data om indvandrede opholdsgrundlag har været publiceret siden 2014 i bl.a. [www.statistikbanken.dk/van8a](http://www.statistikbanken.dk/van8a). De reviderede data erstatter denne opgørelse.

I forbindelse med revisionen af data, er der anvendt en ny imputeringsmetode, der giver et sikrere resultat på individniveau. En del af forbedringerne skyldes, at det er muligt at anvende flere imputeringsvariable, på et lavere aggregeringsniveau. De nye imputeringsvariable er familierelation, oplysning om påbegyndt uddannelse, socio-økonomiske status, og lønmodtageroplysninger. Derudover har det været muligt at anvende enkeltlande, i stedet for landegrupperinger, enkelt aldre, i stedet for aldersgrupperinger, samt indvandringsdato i stedet for indvandringsår.

I de reviderede data, er der færre der har fået koblet en opholdstilladelse fra Udlændingestyrelsen. Det skyldes, at i de tidligere data, blev en opholdstilladelse fra Udlændingestyrelsen anvendt, uanset antal dage mellem tilladelsesdato og indvandringsdato, hvor der i de reviderede data kun anvendes tilladelser fra Udlændingestyrelsen fra tre måneder før og 30 dage efter indvandringen. Samlet for perioden 1997-2019 er der 3 pct. flere i populationen, der har fået en imputeret opholdstilladelse.

På individniveau har 22 pct. fået en anden imputeret tilladelse i de reviderede data i forhold til de tidligere imputeringer, når vi sammenligner på Kategori-niveau. Denne andel er større, hvis opholdsgrundlaget er på et lavere aggregeringsniveau. En relativt stor andel af skift i imputering er sket mellem familiesammenføring og tilladelse til asyl mv., samt mellem arbejde og studier både indenfor og uden for EU.

På det aggregerede niveau i hele perioden er der 2 pct. færre med tilladelse til asyl mv. og 1 pct. flere med tilladelse til erhverv, når der sammenlignes på Kategori-niveau. Der er ikke ændret i antal på de øvrige niveauer.

## **5. Register med opholdsgrundlag for indvandrede personer med seneste indvandring fra 1997**

### **5.1. Hvad hedder registret**

Register med første opholdstilladelse for personer der er indvandret fra 1997, er dannet som moduldata i Personstatistikdatabasen:  
d101200.psd\_indvandrede\_opholdsgrundlag

Data i registeret svarer til de opgørelser der bliver publiceret i [www.statistikbanken.dk/van8a](http://www.statistikbanken.dk/van8a). Det er seneste version der skal anvendes.

### **5.2. Variable**

Variablene er beskrevet i TIMES, hvor der er tilknytning til de relevante kode-tekst-tabeller.

REFERENCETID  
VERSION  
PERSON\_ID  
INDVANDRINGSDATO  
TILLADELSESDATO  
FORKLAR  
KATEGORI  
GRUNDLAG  
IMPUTERET  
STATSB



## 6. Register med første og seneste opholdsgrundlag for indvandrere

### 6.1. Hvad hedder registret

Register med første og seneste opholdstilladelse for indvandrere i befolkningen, er dannet som moduldata i Personstatistikdatabasen:  
d101200.psd\_status\_opholdsgrundlag

### 6.2. Hvordan dannes registret

Registeret opgør indvandrere pr. 1. januar fra 2005 og frem efter første og seneste opholdstilladelse.

For indvandrere der har en indvandring siden 1997, kan der være flere opholdstilladelser, både i forbindelse med genindvandring og i forbindelse med skift af opholdstilladelse ved bopæl i Danmark. Hvis der kun er én tilladelse, vil første og seneste være ens. Den første og seneste opholdstilladelse opgøres på "Kategori", "Grundlag", "Forklar" og "Opholdsgrundlag". "Opholdsgrundlag" er en aggregeret variabel, der kan opgøre både opholdstilladelser for indvandrede før 1997 og efter på en sammenlignelig måde. Disse opholdstilladelser hentes fra d101200.psd\_opholdsgrundlag\_fra1997 efter første og seneste indvandringsdato eller tilladelsesdato, alt efter om tilladelsen er imputeret eller stammer fra Udlændingestyrelsen. Bemærk at Kategori\_seneste, Grundlag\_seneste og Forklar\_seneste er sat til 9997, hvis personen er dansk/nordisk statsborger på referencetidspunktet, og Opholdsgrundlag\_seneste er sat til 999.

For indvandrere med seneste indvandring før 1997, findes kun første opholdstilladelse til og med referencedato 31.12.2018, som dermed vil være ens i både første og seneste opholdstilladelse, med mindre vedkommende er blevet dansk/nordisk statsborger. Fra d101200.psd\_opholdsgrund\_foer1997\_survey hentes disse opholdstilladelser.

### 6.3. Variable

REFERENCETID  
VERSION  
PERSON\_ID

KATEGORI\_FOERSTE  
GRUNDLAG\_FOERSTE  
FORKLAR\_FOERSTE  
OPHOLDSGRUND\_FOERSTE  
IMPUTERINGSSTATUS\_FOERSTE  
TILLADELSESDATO\_FOERSTE  
KATEGORI\_SENESTE  
GRUNDLAG\_SENESTE  
FORKLAR\_SENESTE  
OPHOLDSGRUND\_SENESTE  
IMPUTERINGSSTATUS\_SENESTE  
TILLADELSESDATO\_SENESTE

### 6.4. Værdisæt til opholdsgrundlag

Opholdsgrundlag er en variabel, der er dannet for at kunne sammenligne opholdstilladelser for indvandrede før 1997 med opholdstilladelser som er tildelt fra Udlændingestyrelsen til personer, der er indvandret fra 1997 og frem.

## Værdisæt til Opholdsgrundlag:

Kode	Tekst
100	Asyl mv.
201	Familiesammenført til ægtefælle, Refererer til flygtning
202	Familiesammenført til ægtefælle, Refererer til dansk/nordisk statsborger
203	Familiesammenført til ægtefælle, Refererer til udlænding, men ikke flygtning
204	Familiesammenført til ægtefælle, Uoplyst referenceperson
301	Familiesammenført til forælder, Refererer til flygtning
302	Familiesammenført til forælder, Refererer til dansk/nordisk statsborger
303	Familiesammenført til forælder, Refererer til udlænding, men ikke flygtning
304	Familiesammenført til forælder, Uoplyst referenceperson
401	Familiesammenført til andet familiemedlem, Refererer til dansk/nordisk statsborger
402	Familiesammenført til andet familiemedlem, anden referenceperson
501	Arbejde til EU-borger
502	Arbejde til ikke-EU-borger
601	Studier – herunder praktikanter – til EU-borger
602	Studier – herunder praktikanter – til ikke-EU-borger
701	Au pair
801	Andet
901	Øvrige EU-borgere
999	Nordiske statsborgere

## 7. Kvalitet af register om opholdsgrundlag for indvandrere fra 1997 og frem

### 7.1. Kvalitet

Registret med opholdstilladelser for indvandrede, med seneste indvandring efter 1997 har generelt en høj kvalitet, med de forbehold der er, når en større andel af recordene er imputerede.

Der er høj sammenhæng mellem fx familierelation, arbejdsmarkedsstatus og lønmodtagerstatus i forhold til de typer af opholdstilladelser det kan forventes, både for de der har fået tilknyttet en opholdstilladelse fra Udlændingestyrelsen, og de der har fået en imputeret tilladelse. Fx har 99 pct. af de der har fået en imputering til "Familiesammenføring, Mindreårige børn" også en forældrerelation i CPR-registret, og 93 pct. af de der har fået en imputeret "Familiesammenføring, Ægteskab eller fast samlivsforhold", der har en ægtefællerelation. Dette gælder henholdsvis 97 pct. og 95 pct. af de der har fået tilknyttet en tilsvarende opholdstilladelse fra Udlændingestyrelsen.

Som beskrevet i punkt 3.3 kan det konstateres, at 96 pct. af opholdstilladelserne i registret kan betragtes som korrekte på individniveau. De sidste 4 pct. kan også betragtes som relativt sikre på et overordnet niveau, men der kan typisk være byttet rundt på opholdstilladelser blandt personer med samme karakteristika.

### 7.2. Forbehold

Det anbefales at være meget varsom med at udarbejde analyser og konklusioner baseret på et mindre antal imputerede opholdsgrundlag.

I forbindelse med anvendelse af oplysninger om opholdstilladelser, er det fx vigtigt at være opmærksom på målefejl, der kan være, når man anvender imputerede data.

Hvis man fx vil bruge data til at beskrive nogle karakteristika ved fx flygtningen versus ikke-flygtninge, så kan det gøre en forskel om man inkluderer de imputerede værdier eller ej i sin opgørelse.

**Et tænkt eksempel kan være:**

Vi har to lige store grupper af flygtninge og ikke-flygtninge.

- Alle flygtninge ansættes i kommunalt tilbud som sosu-assistenten.
- Alle ikke-flygtninge ansættes i privat regi som landbrugsmedhjælpere.

Men: Halvdelen af opholdsgrundlagene er imputeret og for 10 pct. af individerne er der byttet om på opholdsgrundlag, så 'sande' flygtninge har værdien ikke-flygtninge og omvendt.

I dette tilfælde vil vi få en fordeling på beskæftigelsestyper, hvor flygtninges beskæftigelse fordeler sig på 90 pct. sosu-assistenten og 10 pct. landbrugsmedhjælpere.

Ikke-flygtninges beskæftigelse fordeler sig på 10 pct. sosu-assistenten og 90 pct. landbrugsmedhjælpere.

I eksemplet her medfører den manglende præcision i opholdsgrundlaget at forskellene i beskæftigelsesfordelingen mellem opholdsgrupperne fremstår mindre end de i virkeligheden er. Eksemplet kan anvendes på alle typer karakteristika, og er en generel problematik, der skal medtages i forbindelse med analyser, hvor data delvist er fra survey eller er imputerede.

**Tabel 1.**  
**Indvandrede 1997-2020 efter alder**

	0-9 år	10-14 år	15-17 år	18-24 år	25-39 år	40+	I alt
<b>ANTAL</b>							
1997	2.092	1.042	1.087	5.318	8.891	2.814	21.244
1998	2.241	1.151	1.066	5.973	9.560	2.535	22.526
1999	1.928	1.003	1.031	6.058	9.152	2.272	21.444
2000	2.608	1.382	1.193	6.687	10.084	2.699	24.653
2001	2.979	1.717	1.385	7.038	10.814	3.041	26.974
2002	2.376	1.347	1.175	7.401	9.465	2.196	23.960
2003	1.738	987	998	7.150	8.374	1.919	21.166
2004	1.452	841	951	7.371	8.808	1.794	21.217
2005	1.282	579	843	8.713	10.113	1.917	23.447
2006	1.500	660	768	9.770	12.262	2.717	27.677
2007	2.118	776	811	11.748	16.500	3.641	35.594
2008	2.824	971	977	14.197	20.252	4.577	43.798
2009	2.401	768	934	14.029	17.824	3.628	39.584
2010	2.465	818	1.044	14.313	19.204	3.529	41.373
2011	2.687	839	1.050	14.692	19.783	3.543	42.594
2012	2.918	803	1.025	15.426	20.506	4.216	44.894
2013	3.392	1.043	1.053	16.277	24.304	5.062	51.131
2014	4.384	1.427	1.415	17.450	28.308	6.601	59.585
2015	6.925	2.137	1.962	19.640	31.653	7.645	69.962
2016	6.270	1.976	1.850	19.499	29.274	6.957	65.826
2017	4.488	1.427	1.523	18.438	28.300	6.901	61.077
2018	3.748	1.218	1.175	18.272	27.934	7.149	59.496
2019	3.243	1.025	1.120	17.532	27.186	6.905	57.011
2020	2.370	905	843	11.778	21.366	6.152	43.414
<b>ANDEL MED IMPUTERET OPHOLDSGRUNDLAG (I PCT.)</b>							
1997	49	78	46	16	15	18	24
1998	36	79	41	7	9	13	17
1999	35	78	39	7	9	14	16
2000	29	83	47	6	8	11	16
2001	18	84	53	6	8	11	16
2002	15	80	50	7	8	12	15
2003	9	71	45	6	9	15	13
2004	17	59	44	9	11	17	14
2005	18	40	32	7	9	17	11
2006	17	28	41	11	12	20	14
2007	14	19	31	13	15	21	15
2008	17	18	32	17	24	34	22
2009	16	18	19	12	16	23	15
2010	20	20	19	12	17	23	16
2011	35	31	28	21	24	33	25
2012	34	33	25	27	26	32	28
2013	25	21	23	26	23	26	25
2014	21	20	17	18	18	22	18
2015	30	30	28	40	16	11	24
2016	6	6	5	6	8	10	7
2017	6	4	4	5	7	9	6
2018	6	5	4	8	8	10	8
2019	6	8	7	8	7	10	8
2020	7	9	8	7	9	11	9

**Tabel 2.**  
**Indvandrede 1997-2020 efter opholdsgrundlag**

	Asyl mv.	Familie- sammen- føring	Studie mv.,				Erhverv	EU/EØS,			Det øvrige opholds- område	I alt
			Uddan- nelse	Au pair	Prakti- kanter	Øvrige grunde		Løn- arbejde	Uddan- nelse	Øvrige grunde		
<b>ANTAL</b>												
1997	3.991	6.710	2.168	413	725	184	1.033	2.329	1.263	1.640	788	21.244
1998	3.056	8.076	2.634	431	792	194	958	2.289	1.539	1.668	889	22.526
1999	2.433	8.059	2.856	537	739	105	936	1.997	1.630	1.263	889	21.444
2000	3.744	8.712	2.963	782	1.120	89	969	1.970	1.770	1.462	1.072	24.653
2001	4.682	9.313	3.168	805	1.246	104	1.021	1.934	1.844	1.456	1.401	26.974
2002	2.732	7.463	3.973	865	1.364	267	1.587	1.683	2.035	1.362	629	23.960
2003	1.870	4.414	5.041	865	1.138	620	1.655	1.675	2.215	1.438	235	21.166
2004	1.329	3.619	3.984	1.068	1.060	838	2.656	1.694	3.154	1.587	228	21.217
2005	1.144	2.813	4.450	1.007	1.460	581	3.760	2.106	3.941	1.993	192	23.447
2006	1.170	2.890	3.781	1.239	1.892	474	6.072	2.895	4.715	2.213	336	27.677
2007	1.271	3.693	4.296	1.627	2.485	543	8.830	3.859	5.442	3.189	359	35.594
2008	1.390	3.267	5.143	2.081	2.378	953	6.715	10.429	6.396	4.666	380	43.798
2009	1.288	3.702	4.546	2.097	1.778	1.125	3.855	8.879	7.456	4.517	341	39.584
2010	2.030	4.032	2.590	2.004	1.328	3.101	4.234	8.430	8.383	4.917	324	41.373
2011	2.307	2.688	4.774	1.902	1.209	735	5.266	9.458	8.328	5.586	341	42.594
2012	2.522	2.742	5.298	1.469	1.132	705	4.834	10.692	8.768	6.438	294	44.894
2013	3.483	4.357	5.806	1.576	1.174	983	5.927	11.969	8.985	6.522	349	51.131
2014	6.056	5.048	6.036	1.478	1.264	910	7.258	14.416	9.228	7.481	410	59.585
2015	10.423	9.432	7.097	1.292	863	1.366	6.812	14.942	8.836	8.486	413	69.962
2016	7.183	7.259	7.687	1.063	1.008	1.871	7.208	15.155	9.328	7.657	407	65.826
2017	2.355	5.779	7.630	1.145	1.527	2.215	6.774	16.030	9.079	7.901	642	61.077
2018	856	3.955	7.920	1.002	1.923	2.729	7.098	16.888	8.545	7.985	595	59.496
2019	660	2.638	7.690	842	1.917	3.846	7.219	15.679	7.997	8.143	380	57.011
2020	434	3.076	3.596	503	1.314	2.164	4.481	13.066	7.232	7.132	416	43.414
<b>ANDEL MED IMPUTERET OPHOLDSGRUNDLAG (I PCT.)</b>												
1997	24	28	18	27	18	30	28	12	13	37	29	24
1998	20	20	5	17	7	21	17	11	9	33	16	17
1999	21	20	5	23	8	7	19	10	5	30	19	16
2000	17	22	5	17	5	12	15	10	4	27	17	16
2001	16	23	6	18	6	10	13	14	7	22	14	16
2002	18	21	9	15	5	10	16	15	3	19	11	15
2003	17	24	5	16	6	10	13	16	5	19	20	13
2004	22	25	6	15	8	13	15	17	6	22	20	14
2005	18	15	6	8	5	9	10	17	8	24	14	11
2006	18	11	9	10	7	14	13	22	11	23	36	14
2007	15	14	14	9	13	20	19	20	9	17	25	15
2008	16	21	13	8	15	26	32	34	10	22	27	22
2009	29	23	10	8	14	23	13	23	6	13	18	15
2010	16	25	16	15	14	16	18	20	7	15	29	16
2011	24	27	20	15	16	29	19	32	18	35	41	25
2012	13	26	20	19	18	29	15	35	27	42	35	28
2013	7	17	22	17	14	22	14	32	30	35	26	25
2014	6	16	10	10	10	16	11	26	21	29	20	18
2015	12	19	33	13	16	27	17	20	44	34	29	24
2016	4	6	5	4	3	10	8	10	7	9	16	7
2017	4	6	6	5	2	11	5	8	4	8	6	6
2018	13	5	9	8	4	18	5	9	7	7	12	8
2019	19	11	9	7	2	12	5	9	7	7	12	8
2020	26	8	6	10	4	26	6	9	6	9	15	9

**Tabel 3.****Indvandrede efter opholdsgrundlag i nye opgørelser i forhold til tidligere opgørelser**

Asyl mv.	Familie- sammen- føring	Studie mv.	Erhverv	EU/EØS	Det øvrige opholds- område	I alt
----------	-------------------------------	------------	---------	--------	----------------------------------	-------

**PROCENT**

1997	96	103	96	110	102	93	100
1998	98	102	96	108	102	92	100
1999	99	101	96	112	102	96	100
2000	95	104	96	105	102	96	100
2001	90	107	97	105	102	96	100
2002	90	105	98	106	101	99	100
2003	97	105	99	100	100	84	100
2004	100	102	98	104	99	90	100
2005	101	99	99	102	101	89	100
2006	101	97	99	102	100	119	100
2007	100	97	101	103	98	105	100
2008	99	96	100	113	97	110	100
2009	109	97	100	97	101	96	100
2010	98	97	101	99	100	101	100
2011	99	95	101	99	101	106	100
2012	99	96	101	97	101	98	100
2013	99	97	101	100	100	93	100
2014	99	98	101	100	100	91	100
2015	100	98	101	102	100	89	100
2016	98	100	101	100	100	103	100
2017	97	99	102	98	100	99	100
2018	92	98	102	99	100	102	100
2019	87	101	102	96	100	96	100