

Datadokumentation MGR-lite

Proveniens	1
Sammenkøring af data	3
Rensning af data	3
CPR-numre.....	3
Håndtering af dubletter.....	4
Tildeling af henvisningsnumre i CPR 1968 og 1969.....	5
Sammenligning af relationer	7
Dannelse af MGR_PNR	7

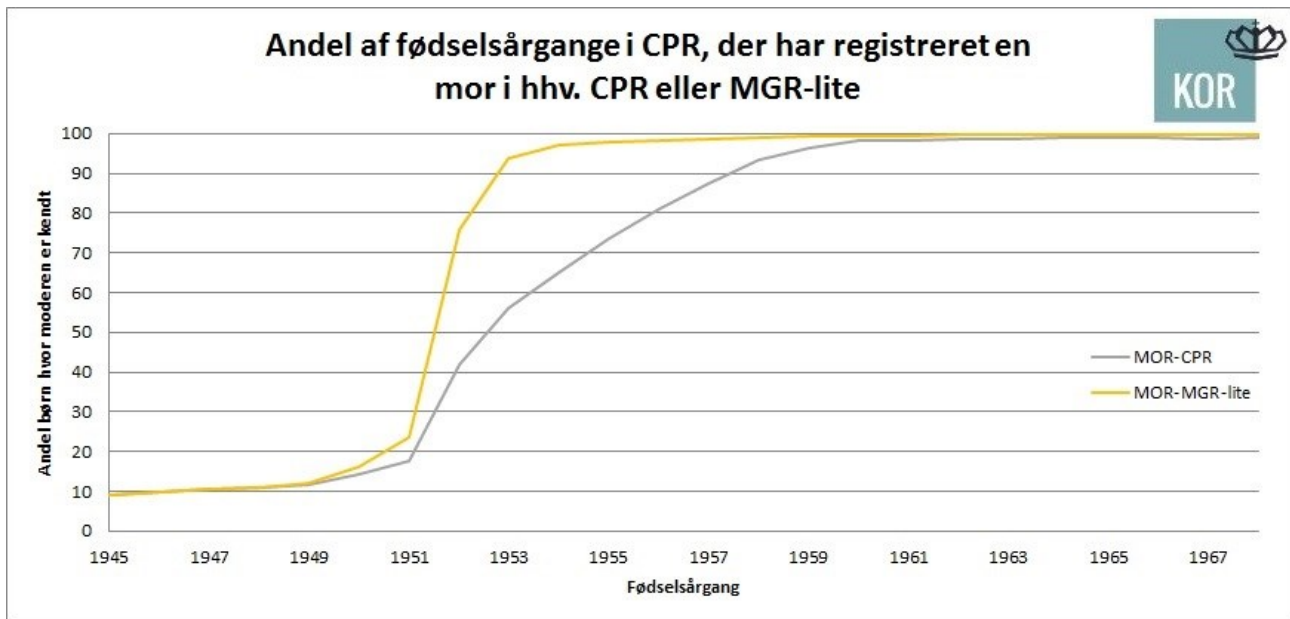
Proveniens

MGR-lite er dannet af Det Koordinerende Organ for Registerforskning til anvendelse i et pilotprojekt, der skulle vise, hvordan man kunne danne et Multigenerationsregister (MGR) over familierelationer for alle danskere født efter 1920.

MGR-lite er dannet af tre kopier af CPR-registret fra hhv. 1968, 1969 og 2013, der er bevaret i Rigsarkivet. De tre versioner er kørt sammen for at kunne danne de bedste familierelationer for personer fra CPR-registret.

Indtil 1978 var der ikke historik på familierelationer i CPR-registret, dette kombineret med at CPR-registret fortsatte praksis fra folkeregisterkortene med at slette relationen mellem børn og forældre, når børnene fyldte 15 år har medført, at relationerne mellem børn født før 1962 og deres forældre er gået tabt i det nuværende CPR-register.

I 2000 tallet lavede Kirkeministeriet et stort arbejde for at genoprette relationer mellem børn født før 1959, hvor de på baggrund af opslag i kirkebøgerne på ny indtastede relationer mellem børn og forældre. Der er bevaret kopier af CPR-registret fra 1968 og frem i Rigsarkivet. Disse kopier indeholder relationer mellem forældre og børn født efter hhv. 1952 og 1953. Disse oplysninger er kørt sammen med en kopi af det nuværende CPR-register (en kopi i Rigsarkivet fra 2013), hvorved der er etableret næsten komplette familierelationer for forældre og børn født efter 1952.



Figur 1: Af grafen fremgår andelen af børn fra fødselsårgangene 1945-1968, hvor der er identificeret en mor i hhv. MGR-lige (gul) og CPR-registret anno 2013 (grå).

Rigsarkivets kopier af CPR-registret er øjeblikksbilleder af data på det tidspunkt, hvor de er overført til arkivet. Det betyder at eventuelle fejl og uregelmæssigheder, der var i data ikke er blevet rettet. F.eks. optræder der tal i flere navne (Mar1anne), individer kan have fået tildelt CPR-numre (PNR) fra to forskellige kommuner, hvis de er flyttet i tiden omkring CPR-registrets oprettelse m.v.

CPR-registret var oprindeligt programmeret i en hierarkisk databasemodel, det betyder f.eks. at familier dannes via såkaldte henvisningsnumre efter følgende systematik:

Familiestatus	PNR	Henvisningsnummer
Mand	1	2
Kvinde	2	1
Barn	3	1
Barn	4	1

Familien dannes altså via henvisningsnummer fra hhv. børn og moderen til faderen. Dette har vi normaliseret til en relationel databasefil, hvor der ud fra hvert individ er angivet PNR på hhv. moder og fader, hvis det er kendt.

Indtil 1978 var familier registreret på basis af boligfællesskab i stedet for biologiske familier. Det vil sige at børn der boede i en plejefamilie var registreret med henvisning til plejefamilien. Dette har vi s'g at rette op på ved at anvende de familierelationer, der er kvalitetssikret af Kirkeministeriet. Men enkelte børn vil have en henvisning til en plejeforælder baseret på boligfællesskabet. Dette vil gøre sig gældende for børn født 1953-1960.

Sammenkøring af data

CPR-registret fra 1968 var så fejlbehæftet (der er f.eks. ikke nogen angivelse af fødesogn), at vi har taget udgangspunkt i en version af CPR-registret fra 1969 og suppleret med de familierelationer for børn født i år 1953, der var blevet slettet i 1969.

Det rensede CPR register fra 1969 er sammenkørt med forælder- og/eller ægtefælle- oplysninger fra CPR-registret fra 1968. Dog således, at de få cpr-numre der findes i 1968 men ikke i 1969 ikke medtages.

Derudover er det rensede 1969 CPR register sammenkørt med forælder- og eller ægtefælle- oplysninger fra en version af CPR-registret fra 2013.

Rensning af data

Rensning er foretaget af Kim Hartung Jørgensen (KJG) fra Danmarks Statistik for KOR i forbindelse med pilotprojekt for Multigenerationsregisteret.

CPR-numre

I 1968 og 1969 CPR registre burde Henvisningsnummeret (HENVNR) være en henvisning fra et individ enten til en ægtefælle eller til en forælder (hjemmeboende børn under 15 år). Dette er ikke altid tilfældet.

De oprindelige HENVNR er ikke med i disse data, men er benyttet til at beregne tildelt PNR for far og mor i 1968 og 1969.

Følgende er en opsummering af den oprensning der er sket:

1. Henvisningsnumrene burde være 10-cifrede som CPR-numrene er, men der er flere eksempler på 9, eller 11 cifrede numre.
2. Der er 804.000 9-cifrede HENVNR, 1.917.618 10-cifrede og 981.020 11-cifrede.
3. 9-cifrede HENVNR: Sættes '0' (nul) foran, er der 547 der ikke findes som PNR for individer i registret:
 - a. KJG har set på det, og kan ikke umiddelbart finde nogen årsag, såsom tastefejl i HENVNR.
 - b. Det kunne se ud som om dem der henvises til simpelthen mangler som individ i CPR-registret (de 547).
 - c. Alle 9-cifrede har fået tilføjet et foranstillet '0'. Gøres ved indlæsningen.
 - d. Disse accepteres herefter alle ved modulus-11 tjek. Hvorpå det konstateres, at det er et validt PNR.
4. 10-cifrede HENVNR: Der er 2906 der ikke findes som PNR for et individ.
 - a. KJG har set på det, og kan ikke umiddelbart finde nogen årsag, såsom tastefejl i HENVNR.
 - b. Det kunne se ud som om de individer, der henvises til, simpelthen mangler i CPR-registret.
 - c. 5 accepteres ikke ved modulus-11 tjek. Disse er fundet og rettes ved indlæsning ved at se på ægtefælle på samme adresse.
5. 11-cifrede HENVNR: Blandt de 981.020 starter 18.937 med '2' og 962.083 med '6'.
 - a. Fjernes '2' eller '6' tallet, findes alle undtagen 121 som PNR for et individ i CPR-registret. Dette er gjort ved indlæsningen.
 - b. Det kunne også her se ud som om dem der henvises til simpelthen mangler i CPR-registret.
 - c. Altså kan første ciffer i de 11-cifrede (2 eller 6) fjernes ved indlæsningen.

- d. Disse accepteres herefter alle ved modul-11 tjek.
6. Efter ændringerne i 3, 4 og 5 er der i alt 3569 HENVNR der ikke findes som PNR i 1969.
 - a. Af disse findes 550 i CPR1968, hvorfor det formodes at de er døde eller udvandret i løbet af 1968.
 - b. Der er en del personer der er registreret med 2 forskellige PNR.
 - c. De er typisk næsten ens, men med forskellig *STATUS-kode* (ikke med her), f.eks. 0 og 6. Nogle har lidt forskellig fødselsdato, men ellers ens data.
 - d. Ca. 10 % af disse findes typisk med et PNR med samme første 6 cifre, men et andet løbnummer.
7. Der er 668 hvor HENVNR er personens eget. KJG: Disse fjernes (altså HENVNR) ved indlæsning.
8. Der er desuden mindst 11-12.000 hvor henvisningen mellem ægtefæller kun går den ene vej. KJG: Dette rettes der også op på ved indlæsning. For detaljer se SQL-scripts og deres dokumentation.

Samtlige CPR numre i MGR-lite er accepteret ved modul-11 tjek.

Håndtering af dubletter

MGR-Lite tager udgangspunkt i de data som CPR afleverede til Rigsarkivet i 1969, hvor der er foretaget et antal rensninger af data og fjernet nogle historiske (primært forkerte) CPR-registreringer.

Følgende problematikker er håndteret:

1. Der er i alt registreret 131.779 inaktive PNR (STATUS = 5,6,7,8, og 9)
2. Af disse er 362 forsvundne, 31.110 udvandret, 58.217 døde (i 1969).
3. 37.986 er 'gennummereret' (STATUS 6):
 - a. Disse er typisk personer hvor fødselsdagen har været lidt forkert i den oprindelige registrering fra 1968. Der er dog også nogle få, hvor fødselsdatoen ikke er ændret i det nye PNR. Disse personer har så fået et nyt PNR og det gamle markeret 6.
 - b. Det ser ud til, at HENVNR for disse altid (har tjekket ved omfattende stikprøver) peger på det nye PNR personen har fået.
 - c. Børn peger ofte på disse (nu historiske), hvorfor man kun kan finde forælder referencen gennem disse. Dette er rensset ved indlæsning og er beskrevet detaljeret i dokumentationen til sql-scripts.
4. 4.104 er 'nynummereret' (STATUS 5):
 - a. Disse personer er efter al sandsynlighed flyttet (mellem 2 kommuner), men er oprettet med PNR i begge kommuner. Da (hvis) man blev opmærksom på det gav man cpr nummeret i fraflytningskommunen STATUS=5.
 - b. Dette understøttes af, at disse alle har samme fødselsdato.
 - c. Det understøttes også af, at f.eks. fornavn kan være lidt forskelligt, f.eks. 'KIM HARTUNG' i det ene tilfælde og 'KIM H.' i det andet. Dette er rensset ved indlæsning og er beskrevet detaljeret i dokumentationen til sql-scripts.
5. Registreringen af gennummereret og nynummereret har formentlig været mangelfuld.
 - a. I forbindelse med optællingen på fødselsår (1969, 1968 ...) kan det konstateres, at selvom alle gennummererede og nynummererede fjernes (status = 6 og 5), så er der stadig en del dubletter

med forskellige PNR, men samme fødselsdag og navn eller adresse (det er altså personer der optræder 2 gange).

- b. Det ser ud som om disse, ud over at have samme fødselsdag ofte har de samme data på stort set samtlige variable, bortset fra 1 eller 2 variable (dog altid forskellig AENDRDTO). Det er altså samme person.
- c. Her har vi overvejet, at udelade den med ældst AENDRDTO fra optællingen. Dette er Pt. IKKE gjort, da det ikke entydigt er det rigtige at gøre, hvorfor det kræver yderligere analyse.
- d. En senere samkøring med 2013 data, kan formentlig afsløre hvilket PNR der faktisk 'overlevede'. Her er Pt. ikke gjort noget. Ved samkøringen med CPR2013 vil der for disse bare ikke være nogle data for 2013. Dette viste sig ikke at være så simpelt og kræver også mere analyse.

Vi har fjernet/konsolidereret dubletter hvor personer har fået 2 (eller flere) personnumre fordi de er flyttet i tiden omkring oprettelse af CPR, eller har fået et forkert PNR pga. forkert fødselsdato eller køn.

Vi havde overvejet at fjerne de yderligere dubletter vi ved der er, hvor STATUS-koden oftest ikke er markeret som 5 eller 6, men det har vist sig mere kompliceret at finde dubletterne end forventet. Når de så er fundet er det heller ikke trivielt at finde ud af hvilken der skal slettes, samt genoprette forælder og ægtefælle-relationer. Da der vurderes kun at være 5-700 dubletter har vi valgt ikke at fjerne yderligere dubletter.

Tildeling af henvisningsnumre i CPR 1968 og 1969

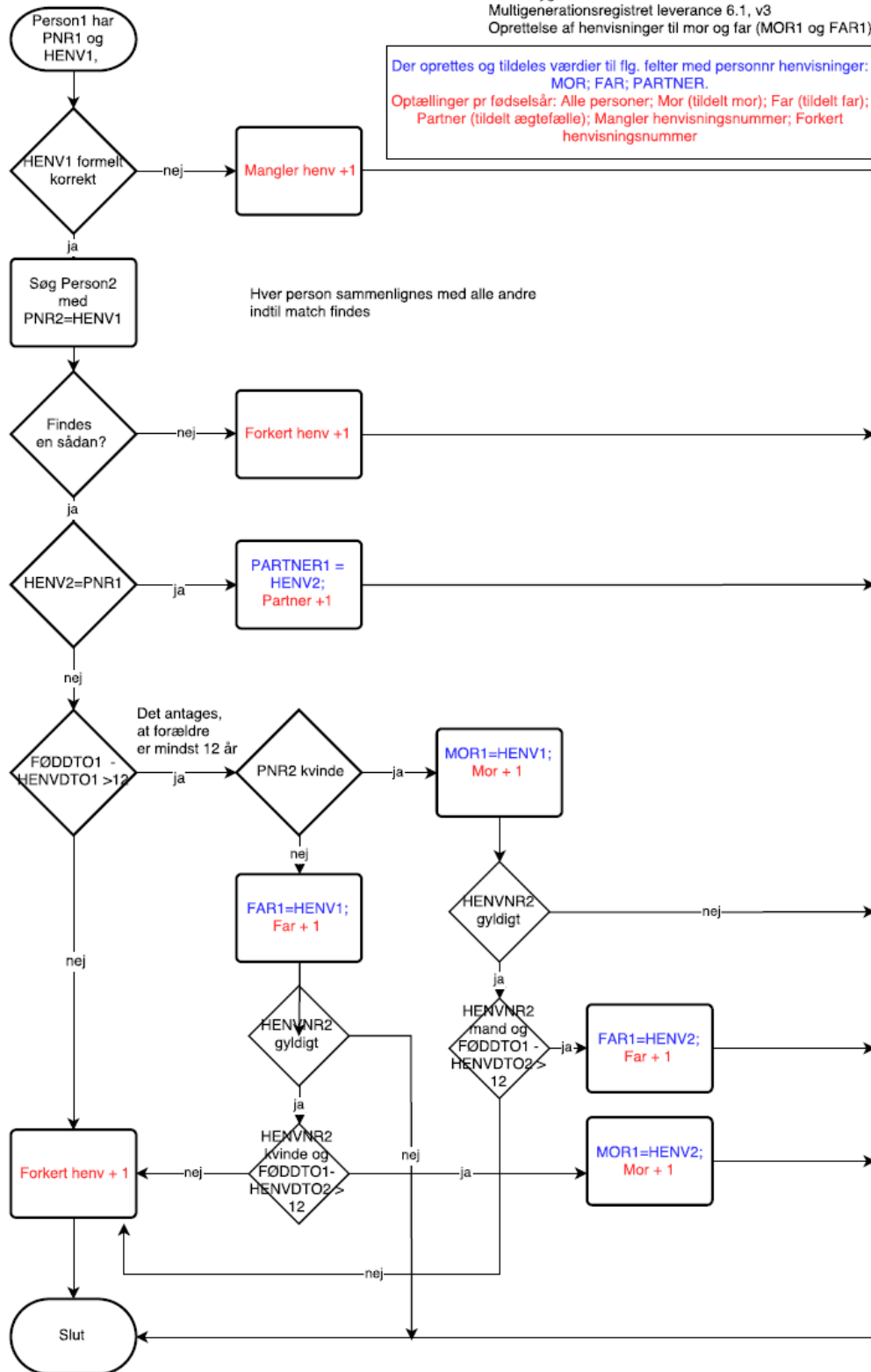
Man tildeler henvisninger til mor (MOR) og far (FAR) vha. CPR1968 og CPR1969 som vist i Fig. 2.

Henvisningsnummeret burde pege på enten en ægtefældes eller en forælders PNR. I praksis er der alle mulige fejl og mangler, som der er rettet op på i rensningen så godt som muligt.

Slutresultatet er at alle ægtefæller peger korrekt på hinanden samt at alle børn peger på den korrekte (ikke historiske) version af både faderen og moderen.

For at se detaljerne for rensningerne se evt. kommentarerne i starten af vedlagte scripts.

Lars Thygesen 13.08.2016
 Multigenerationsregistret leverance 6.1, v3
 Oprettelse af henvisninger til mor og far (MOR1 og FAR1)



Figur 2: Beslutningsdiagram for dannelse af forældre-henvisninger i CPR 1968-1969

Sammenligning af relationer

Familierelationerne for enkelte individer varierer lidt mellem de tre forskellige versioner af CPR, og hver version indeholder relationer mellem børn og forældre, som ikke kan findes i de andre versioner.

Der er derfor gennemført en sammenligning mellem henvisningsnumre til mor og far fra de 3 versioner af CPR. Resultatet er opsummeret i tabel 1 og 2 for de "børn", fra de ti sogne, der fungerede som stikprøve i pilotprojekt for MGR. Resultatet vurderes at være repræsentativt for hele CPR-populationen.

Tabel 1. CPR-mødres variation: Hvor mange børn er registreret med samme eller forskellige mor i de 3 CPR-versioner

Tabel 1. Hvordan passer CPR-mødre sammen

Population: Alle KBF børn matchet med CPR

Antal	alle er ens			to er ens			alle/begge forskellige	
	1968=1969=2013	1968=1969#2013	1968=2013#1969	1969=2013#1968	1968#1969#2013			
Mor 1968 findes Mor 1969 findes Mor 2013 findes	3313	3251	30	1	31	0		
Mor 1968 findes Mor 1969 findes Mor 2013 findes ikke	797		788			1968#1969		
Mor 1968 findes Mor 1969 findes ikke Mor 2013 findes	144			1968=2013		1968#2013		
Mor 1968 findes Mor 1969 findes ikke Mor 2013 findes ikke	100			139		5		
Mor 1968 findes ikke Mor 1969 findes Mor 2013 findes	38				1969=2013	1969#2013	2	
I alt	4392							

Tabel 2. CPR-fædres variation: Hvor mange børn er registreret med samme eller forskellige far i de 3 CPR-versioner

Population: Alle KBF børn matchet med CPR

Antal	alle er ens			to er ens			alle/begge forskellige	
	1968=1969=2013	1968=1969#2013	1968=2013#1969	1969=2013#1968	1968#1969#2013			
Far 1968 findes Far 1969 findes Far 2013 findes	3086	3012	53	1	20	0		
Far 1968 findes Far 1969 findes Far 2013 findes ikke	786		778			1968#1969		
Far 1968 findes Far 1969 findes ikke Far 2013 findes	128			1968=2013		1968#2013		
Far 1968 findes Far 1969 findes ikke Far 2013 findes ikke	86			126		2		
Far 1968 findes ikke Far 1969 findes Far 2013 findes	42				1969=2013	1969#2013	5	
I alt	4128							

Dannelse af MGR_PNR

For at lette anvendelsen af MGR-lite har vi dannet den bedste version af relationen fra et individ til forældrene. Dette er gjort efter følgende regel:

Nyeste registrering af forældrenes CPR nummer trumfer en ældre forældre registrering. Altså hvis der er PNR for en forælder i 2013 vælges denne, hvis der ikke er et forældre PNR i 2013 men et i 1969, så vælges PNR for forældre i 1969. Hvis der ikke er CPR numre for forældre i hverken 2013 eller 1969 men i 1968, så vælges forældres CPR nummer fra 1968.

Det sikres, at de personer, der henvises til som forældre, også findes i CPR, evt. som inaktive (døde, udvandrede, forsvundne, gen- eller nymummerede); i modsat fald fjernes henvisningen.

Det er et fåtal af individer, der har relationer til forskellige forældres CPR-numre i de forskellige versioner af CPR-registret, der er arbejdet med. Se tabel 1 og 2.

MGR-lite indeholder alle CPR numre på personer, der har været angivet som forældre til et individ i CPR-registret, så det vil være muligt at lave andre regler, end den der er beskrevet ovenfor.

Den valgte fremgangsmåde har forsøgt at tage hensyn til:

- At der er blevet rettet fejl i forbindelse med administrativ brug af CPR-registret. F.eks. kan et fejlagtigt henvisningsnummer være blevet opdaget og korrigeret i forbindelse med melding af flytning på Folkeregistret.
- I forbindelse med indføring af Den Nye Kirkebog i 00'erne foretog Kirkeministeriet en kvalitetssikring og genopretning af relationer mellem forældre og børn, for børn født >1959.
- I 1968 og 1969 blev der registreret henvisningsnummer til plejeforældre (hvis barnet boede der) i stedet for den biologiske familie. Flere af disse vil være blevet rettet med overgangen til Den Nye Kirkebog.
- Kirkeministeriets kvalitetssikring blev foretaget ved opslag og indtastning fra kirkebøgerne. I kirkebøgerne har man registreret den familiære relation mellem forældre og barn, ikke husstandsfællesskab.

Jeppe Klok Due
Specialkonsulent
KOR