

Dato: 17. marts 2015

Grundlag for KORs pilotprojekt for et multigenerationsregister

I rapporten "Registerforskning - nye muligheder nye udfordringer" fra 2013 anbefales det, at der etableres et multigenerationsregister (MGR).

<http://www.registerforskning.dk/wp-content/uploads/sites/16/2014/12/registerforskning2013.pdf>

KOR har haft konsulent Lars Thygesen til at undersøge mulighederne for at gå videre med arbejdet. På baggrund af dette arbejde har KOR januar 2016 besluttet nedenstående projektgrundlag for gennemførelsen af et pilotprojekt.

Baggrund

CPR-registret indeholder information om alle personer bosiddende i Danmark fra 1968 og frem. Registret har i samspil med en lang række andre registre, der benytter samme identifikation, vist sin værdi i forbindelse med talrige store befolkningsundersøgelser af årsagerne til ophobning af sygdomme i familier, samt al slags forskning, hvor familier og generationsforhold spiller en rolle.

For de fleste personer født i Danmark 1960 eller senere indeholder CPR- registret oplysning om identiteten af mor og i langt de fleste tilfælde også far. For personer født før 1960 er denne information imidlertid kun tilgængelig i meget få tilfælde. Blandt alle personer født i Danmark og som er registreret i CPR, er der over to millioner født 1920-1959, hvortil der ikke er oplysning om identiteten af mor. Konsekvensen er, at man for langt de fleste sygdomme, kun kan beskrive den familiære ophobning af sygdomme eller hændelser inden for to generationer.

En udvidelse af generationsoplysningerne i CPR-registeret vil give dansk forskning en unik mulighed for at studere familiære ophobning af sygdomme eller andre hændelser indenfor 3 eller flere generationer. Muligheden for at koble befolkningen sammen i familier kan blive et afgørende løft i de unikke muligheder vi har for befolkningsbaseret forskning.

KOR

Rigsarkivet
Rigsdagsgården 9
1218 København K

Telefon: 4171 7211

Mail: jkd@sa.dk

Web: registerforskning.dk

Sagsnr.: 16/0601

Formål med projektet

At lave et pilotprojekt, der fokuserer på at etablere mere komplette familierelationer for den population, der eksisterer i CPR-registret. Dette skal i pilotprojektet gøres for en repræsentativ stikprøve af sogne, der kan fungere som et proof of concept.

At kunne foreslå udførelse, finansiering og organisering af projektet i fuld skala.

At anbefale muligheder for at udvide registret med personer, der døde før 1968, herunder skal der i praksis laves eksperimenter med at inkludere individer fra dødsårsagsregistret.

Organisation

Styregruppe Professor cand.med. Kirsten Ohm Kyvik, SDU (formand)

Professor, mag. scient. soc. Lisbeth B. Knudsen, AAU og Klinisk professor, cand.med. Merete Osler, KU

Koordinator: Jeppe Klok Due

Sparringsgruppe: Relevante forskere og statistikere, der arbejder med registre og generationsforskning.

Projektleder: Lars Thygesen

Projektlederen leverer månedligt statusrapport (fremdrift og økonomi) til Styregruppeformanden. Der afholdes månedlig statussamtale mellem projektleder, styregruppeformand og koordinator.

Datagrundlag

- CPR-registret (fuldt register)
- Digitaliserede kirkebøger over fødte 1920-1960 fra ti sogne. Der udvælges en stikprøve bestående af repræsentative sogne under hensyn til forskelle i bebyggelse, geografisk placering og befolkningsstørrelse.
- Digitaliserede folketællinger 1930 – 1960 for to af de ti sogne (hhv. land og by)
- Digitaliserede kirkebøger over døde 1920-1942 samme som folketællingerne
- Dødsårsagsregistret 1943-1968
- Folke- og Boligtællingen 1970
- Eventuelt oplysninger om vielser fra de samme sogne

Formålet er at have en tilstrækkeligt stor population til at lave et proof of concept for et projekt, der etablerer familierelationer for personer i CPR-

registret, samt til at sammenligne styrken ved enten at bruge kirkebøger eller folketællinger til kobling af personer over tid.

Endvidere skal der ud fra stikprøverne kunne identificeres eventuelle problemstillinger, der skyldes varians i kildematerialet.

Metode

Undersøgelsen skal afprøve forskellige metoder og kilder for at kunne estimere omkostninger og forventet sandsynlighed for at kunne identificere familierelationerne i et MGR i fuld skala.

Alle de blandt ovenstående materialer, der ikke allerede er digitaliserede, skal digitaliseres, renses og kodes for at muliggøre kobling på tværs af kilderne.

CPR kobles med fødte for de udvalgte ti sogne. De fødte og forældrene skal matches op mod personer i CPR-registret, for at identificere deres cpr-nummer. Herudover skal det forsøges at komplettere registreringen vedr. døde inden 1968 med oplysninger fra dødsårsagsregistret.

CPR kobles med folketællinger, fødte og døde for de udvalgte to sogne for at fastslå familierelationerne.

Testmiljø

Der oprettes et testmiljø hos Danmarks Statistik (DST). Her skal data behandles på en sikker måde, brugernes aktiviteter kan logges og DSTs datamanagere kan hjælpe med at gøre data analyseklar og installere programmet til kobling m.v.

Budget

Det samlede budget må ikke overstige 2,8 mio. DKK.

Indtastning af data	1 mio. DKK
Projektleder 2016	0,3 mio. DKK
Programmør	0,4 mio. DKK
Statistikere	0,2 mio. DKK
DST (datamanagement m.v.)	0,2 mio. DKK
Uforudsete udgifter	0,7 mio. DKK

Produkter

- Digitaliseret datamateriale fra Kirkebøger over fødte (1920-1960) fra ca. ti repræsentative sogne

Kirkebøger over døde (1920-1943) fra to af de ti sogne.

Folketællingerne fra (1920-1960) fra de to sogne

- Testmiljø hos DST, hvor relevant kildemateriale og programmel er installeret
- Database med familierelationer over individer fra de 10 test-sogne, der var levende i 1968, (deres registrering i CPR skal linkes til deres fødselsregistrering (som enten barn eller forælder) i kirkebøgerne)
- Analyse af bortfald (født i andre sogne, udvandrede, døde etc.)
- Evidensbaseret estimat over hvor mange procent af individerne i CPR, der kan genfindes i alene kirkebøgerne fordelt efter kvalitetskode for sammenkoblingen.
- Evidensbaseret estimat over hvor mange procent af individerne i CPR, der kan genfindes i kirkebøgerne og folketællingerne fordelt efter kvalitetskode for sammenkoblingen.
- Evidensbaseret estimat over hvor mange procent af individerne der er døde 1943-1968, der kan findes både i dødsårsagsregistret og kirkebøgerne.
- Sammenligning af forventet styrke af kobling af familierelationer via oplysninger fra hhv. Kirkebøger og Folketællinger + kirkebøger.
- Kodelister til navne mv. samt strategi for datavask af navnefeltet m.v.
- Udarbejdelse af relevant dokumentation.
- Vurdering af kildematerialet (fx kvaliteten af registrering af familierelationer i CPR 1960-1968).
- Anbefalinger til hvordan projektet kan udføres i fuld skala med hensyn til:
 - o Omkostninger
 - o Tid
 - o Drift
 - o Funding
- Udarbejdelse af use cases
- Anbefale muligheder for og omkostninger ved at supplere lakuner (adoptionssager, udvandrede m.fl.) efterfølgende.
- Oversigt over ældre relevante registre, der kan anvendes sammen med MGR
- Forslag til hvordan MGR kan føres længere tilbage i tid.

Udarbejdet af KORs styregruppe for Multigenerationsregisteret marts 2016

Professor, cand.med. Kirsten Ohm Kyvik, SDU (formand)

Professor, mag. scient. soc. Lisbeth B. Knudsen, AAU

Klinisk professor, cand.med. Merete Osler, KU

Specialkonsulent, cand.mag. Jeppe Klok Due, KOR (Projektkoordinator)