

Introduktion til forskningsudtræk fra Dansk Lunge Cancer Register - DLCR

*** LÆS DETTE FØR DU AFKRYDSER VARIABELLISTE TIL DIT FORSKNINGSUDTRÆK ***

DLCR omfatter data fra de deltagende afdelinger, Landspatientregisteret (LPR), Patologiregisteret og CPR registeret til etablering af det samlede udrednings- og behandlingsforløb for patienter diagnosticeret med lungekræft i Danmark.

DLCR indsamler data på basis af den såkaldte DNKK model (Den Nationale Kliniske Kræftdatabase), hvor der er etableret en specifik algoritme for høst af data fra LPR og Patologiregisteret for patienter med en ICD10 diagnosekode svarende til Kræft i luftrøret (DC33*) og Kræft i bronkier og lunge (DC34*). Se DLCR Årsrapporter for en yderligere uddybning af DNKK modellen: <https://www.lungecancer.dk/arbejdsgrupper/dlcr/opbygning/>

DLCR er en webbaseret indtastningsdatabase, hvor udvalgte data for et patientforløb høstes fra de centrale registre og importeres ind i databasen, hvorefter data suppleres med data fra de deltagende afdelinger. For en lang række af de variable, som indsamles fra de centrale registre er der mulighed for klinisk validering foretaget af klinikere ved de indberettende afdelinger.

DLCR består af tre formularer (datasæt): Udredning, Kirurgi og Onkologi. På hver formular registreres et antal variable, som er relevante for beskrivelsen og registreringen af den udredning og behandling, som patienten modtager på sygehusniveau. En patient kan have flere kirurgi- og onkologiformularer, men kun en udredningsformular, svarende til et udredningsforløb for lungekræft diagnosen.

For hver variabel på variabellisten er der angivet hvilken formular - eller datasæt - den pågældende variabel kommer fra: Udredningsformularen (*Udr*), Kirurgiformularen (*Kir*), Onkologiformularen (*Onk*), Activity datasættet (*Act*), eller datasættet for de patientrapporterede variable (*PRO*).

Enkelte variable er registreret på flere formularer, fx T, N, og M: Det kliniske TNM er registreret på Udredning og Onkologi formularen mens det patologiske TNM er registreret på Kirurgi formularen. Man skal derfor være opmærksom på hvilken formular man henter sit TNM fra.

Tre typer variable i DLCR

Der er tre typer variable i DLCR:

- 1) LPR / Patologi variabel **med** klinisk validering
- 2) LPR / Patologi variabel **uden** klinisk validering
- 3) Manuelt indtastede variable, som ikke kan høstes fra de centrale registre: Landspatientregisteret, Patologiregisteret og CPR registeret.

For LPR eller Patologivariable med klinisk validering (jf. Pkt. 1) er der tre versioner af samme variabel, og hver variabel er angivet med sit eget suffix efter variabelnavnet. Eksempler:

Variabel navn	Beskrivelse - tekst
<i>starttid</i>	DNKK genereret startdato for udredning (rå dato fra LPR / DNKK variabel)
<i>starttid_korr</i>	Kliniker korrigeret startdato for udredning (klinisk korrigeret / valideret LPR dato)
<i>starttid_res</i>	Startdato for udredning (resultat felt) (resulterende dato, dvs. den korrigerede dato, hvis datoen fra LPR ikke var den korrekte. Hvis datoen fra LPR var den korrekte, er det denne dato, som ses i denne variabel).
<i>sideangivelse</i>	Sideangivelse (DNKK genereret) (udfald på baggrund af LPR / DNKK variabel)
<i>sideangivelse_korr</i>	Sideangivelse (kliniker korrigeret) (klinisk korrigeret / valideret LPR udfald)
<i>sideangivelse_res</i>	Sideangivelse (resultatfelt) (resulterende udfald, dvs. det korrigerede udfald, hvis udfaldet fra LPR ikke var det korrekte. Ellers er udfald fra LPR angivet).

Til analyser og forskningsudtræk vil det altid være **_res variabelen**, som skal anvendes, da det er den variabel, som indeholder det (i nogle tilfælde) validerede udfald og dermed den klinisk mest valide registrering. I forskningsudtræk fra DLCR udleveres derfor kun **_res variabelen**, hvis der er tale om en såkaldt DNKK variabel med tre versioner. Ved download af variabelliste fra SundK her <https://www.sundk.dk/teknisk-dokumentation/data-til-forskning/> fås en komplet variabelliste over alle variable fra DLCR.

Udover de variable, som findes i en af de tre formularer i DLCR indtastningsdatabasen (Udredning, Kirurgi, Onkologi), er der et stort antal afledte variable inkluderet i DLCR's populationsdatasæt. Det er variable, som dels beskriver vigtige datoer i udrednings- eller behandlingsforløbet eller som fx markerer at dette er patientens første behandling generelt, første kirurgi eller første onkologiske behandling. De afledte variable er alle afledt af andre variable i populationsdatasættet, som er baseret på data og registreringer i LPR eller Patologiregisteret.

Struktur på udleveret datasæt til forskningsbrug

Når SundK trækker data ud til forskningsbrug afgrænses der ofte på diagnosedato, men også andre afgrænsninger kan være relevante, afhængigt af det enkelte forskningsprojekt. Dette skal specificeres klart af den ansøgende forsker.

Et DLCR datasæt udleveret til forskningsbrug fra SundK omfatter flere observationer/rækker per patient. Data for hver af de tre formularer i DLCR udgør en række, svarende til en række for (data fra) udredningsformularen, en række for (eventuel) kirurgiformular og en række for (eventuel) onkologiformular. En patient kan have flere onkologiformularer, hvis der fx er givet både neo-adjuverende og adjuverende onkologisk behandling. Som forsker skal du forvente flere rækker per patient i det udleverede datasæt, og ikke at datasættet er struktureret med en række per patient.

Som udgangspunkt er datasættet organiseret således, at afledte variable kommer først, dernæst variable for de tre formularer, som er lagt lige efter hinanden med variable fra formular for Udredning, Onkologi og Kirurgi.

Dokumentation af tre datasæt for DLCR

SundK har dokumenteret i alt tre datasæt for DLCR og to datasæt for Mesotheliom databasen:

1) DLCR's Populationsdatasæt, som består af variable fra de tre formularer Udredning, Kirurgi og Onkologi, samt deraf afledte variable (*Datasætnavn: DLCR_POPULATIONS_DS*)

2) Et datasæt, der omfatter alle udrednings- og behandlingsprocedurer registreret i LPR per patient (*Datasætnavn: DLCR_2_SUPPL_AKTIVITET*)

3) Et datasæt, som indeholder data på PROM data, der blev indsamlet i perioden 2013-2016 (*Datasætnavn: DLCR_3_PRODATA*)

4) DLCR_4_MESOTHELIOM_POP og DLCR_5_MESOTHELIOM_SUPPL_AK er de to datasæt for Mesotheliom databasen under DLCR.

Datasættet *DLCR_POPULATIONS_DS* danner altid grundlag for forskningsudtræk fra DLCR. Hvis formålet for et givent forskningsprojekt kræver adgang til øvrige udrednings- og/eller behandlingsprocedurer for den enkelte patient kan der være behov for udtræk fra datasættet *DLCR_2_SUPPL_AKTIVITET*. Derfor skal variable i dette datasæt også afkrydses, hvis der er brug for udrednings- og/eller behandlingsprocedurekoder fra LPR udover hvad der findes i datasættet *DLCR_POPULATIONS_DS*.

På SundK's hjemmeside er en komplet dokumentation af DLCR og DLCR's datasæt, hvor alle variable og udfaldsrum er dokumenteret for databasen. Dokumentationen kan downloades til egen PC.

Læs mere om DLCR her: <https://www.sundk.dk/kliniske-kvalitetsdatabaser/dansk-lunge-cancer-register-dlcr/>

DLCR's dokumentation kan tilgås her: <https://www.db-dokumentation-sundk.dk/Public/Databases.aspx?db2=1000000449>

Udarbejdet af:

Henriette Engberg

Epidemiolog for DLCR, SundK