

Vejledning i udtræk fra Landspatientregisteret

Hvilken type data indeholder Landspatientregisteret (LPR)? Hvordan er registeret bygget op? Hvordan udarbejder du en udtræksspecifikation fra registeret, og hvad skal du særligt tage højde for i denne forbindelse?

Denne vejledning henvender sig til forskere og andre som - typisk via Sundhedsdatastyrelsens Forskerservice - skal bruge data fra LPR (1977-2019). Vejledningen afspejler, hvad der typisk ønskes i et dataudtræk fra LPR. For en yderligere forklaring af variable og registreringspraksis kan vi henvise til dokumentationen af LPR og Fællesindholdet for basisregistrering af sygehuspatienter.

Dette dokument henviser til den somatiske del af LPR. LPR indeholder dog også psykiatriske kontakter fra og med 1995. Ved udtræk af psykiatriske kontakter skal du angive, at du ansøger om data fra LPR-PSYK. LPR-PSYK er med få undtagelser opbygget på samme måde som det somatiske LPR.

Vejledningen omfatter ikke den forskerrettede LPR3-datamodel (LPR3_F).

Hvad indeholder LPR?

LPR er et nationalt sundhedsregister, der omfatter indberetning af alle patienters kontakter til offentlige og private somatiske og psykiatriske sygehuse. Det vil sige, hver gang en person har været i kontakt med det danske sygehusvæsen – for eksempel i forbindelse med en indlæggelse – registreres en række oplysninger om patientkontakten.

Ved en patientkontakt forstås:

- > En indlæggelse
- > En ambulante kontakt (besøg og ydelser)

For alle kontakter er der registreringer af rent administrativ karakter, blandt andet hvornår indlæggelsen fandt sted, på hvilket sygehus og hvilken afdeling. Derudover er der oplysninger om

aktionsdiagnose (og eventuelle relevante bidiagnoser) for kontakten samt eventuelle undersøgelser, behandlinger og operationer (procedurer) registreret på kontakten.

Overblik over oplysninger i LPR:

- › Bopælskommune
- › Sygehus og afdeling
- › Indlæggelses- og udskrivningstidspunkt
- › Diagnoser
- › Behandlinger og operationer

Registerets omfang og complethed af data fra sygehusenes patientregistreringer gør LPR-data-brugbare i forbindelse med forskning inden for sundhedsområdet.

Periode

De somatiske oplysninger i LPR dækker perioden fra 1977 og fremefter. Patientkontakter ved psykiatriske sygehuse er registreret siden 1995. Skadestue- og ambulante kontakter er registreret siden 1994, hvor skadestuekontakter blev erstattet af akutte ambulante kontakter i 2014.

Hvad indeholder LPR *ikke*?

Vi får ofte forespørgsler om oplysninger, som LPR ikke indeholder. Her svarer vi på de hyppigst forekommende:

LPR indeholder ikke:

- › Patientoplysninger fra praksissektoren, eksempelvis alment praktiserende læge (data findes delvist i Sygesikringsregisteret)
- › Oplysninger om personers generelle medicinforbrug (data om lægemidler købt på recept findes i Lægemiddelstatistikregisteret)
- › Oplysninger om personer er levende eller døde (data findes i CPR-registeret)
- › Oplysninger om dødsårsager (data findes i Dødsårsagsregisteret)
- › Fuldstændigt journalindhold. Det er kun udvalgte oplysninger om en sygehuskontakt, som indberettes til LPR.

LPR er kontaktbaseret

En patientkontakt starter med en indskrivningsdato - typisk en patients indlæggelsesdato – og afsluttes med en udskrivningsdato. Indberetningen til LPR sker på baggrund af registreringerne ved kontaktens afslutning. Der skelnes mellem uafsluttede ambulante og stationære patientkontakter. Hver kontakt er i registeret tildelt et recordnummer, og alle data (tabelrækker) med

samme recordnummer rummer oplysninger om denne unikke kontakt (se mere nedenfor om hovedtabel og satellittabeller i LPR).

LPR er baseret på enkeltkontakter. Det betyder, at der ikke direkte er information om længerevarende patientforløb i registeret. Eksempelvis vil et patientforløb med et ambulante besøg, en indlæggelse og efterfølgende ambulante besøg udgøre i alt tre kontakter i LPR.

Forskerservice videregiver data fra den del af LPR, der opdateres månedligt (omkring den 10. i hver måned). Det er muligt at tilgå data om både afsluttede og uafsluttede kontakter.

Et levende register

LPR opdateres løbende med indberetninger af patientkontakter. Indberetninger om kontakter kan indløbe lidt forsinket, eller der kan være rettelser til ældre indberetninger.

Det betyder, at to udtræk baseret på samme aftale om dataadgang, men foretaget på forskellige tidspunkter, kan give lidt forskellige dataudtræk.

Desuden ændrer recordnummeret sig i forbindelse med rettelser, hvorfor man ikke kan koble nyere og ældre udtræk via recordnummeret. Hvis du på et senere tidspunkt ønsker at udtrække de samme kontakter fra LPR, skal du benytte kontaktnøglen, der udgøres af variablene: v_cpr, c_sgh, c_afd, c_patttype, d_inddto, v_indtime og v_indminut.

Data fra perioden 1977-2004 opdateres ikke. Forskerservice videregiver aktuelt opdaterede udtræk fra og med år 2005.

Offentlige og private sygehuse

LPR rummer oplysninger fra både private og offentlige danske sygehuse. Hvis ikke andet angives i ansøgningen, vil et dataudtræk rumme oplysninger fra såvel offentlige som private somatiske sygehuse.

Afgrænsninger ved udtræk

Det er et krav at følgende oplysninger indgår i udtræksbeskrivelsen for dataudtrækket fra LPR:

- Hvilken tidsmæssig afgrænsning (for eksempel periode fra indlæggelsesdato til udskrivningsdato for afsluttede kontakter)?
- Hvilke kontakter er relevante (udtrækskriterier)
- Skal uafsluttede kontakter inkluderes (kontakttype)?
- Hvilke variable skal udtrækkes (oplysninger)?

Tidsafgrænsning af udtræk

LPR dækker perioden fra 1977 og fremefter. Hvis der er tale om et tidsafgrænset studie, har du ikke behov for et fuldt udtræk fra LPR. Du skal derfor angive de årgange eller den periode, som udtrækket skal dække.

Et udtræk fra LPR kan tidsmæssigt afgrænses på forskellige måder. Oftest vil man benytte sig af kontaktens indskrivningsdato og/eller udskrivningsdato på én af følgende måder:

- Alle aktive kontakter i en periode (afgrænsning på indskrivningsdato og/eller udskrivningsdato)
- Alle kontakter indskrevet i en periode (afgrænsning alene på indskrivningsdato)
- Alle kontakter afsluttet i en periode (afgrænsning alene på udskrivningsdato)

Disse muligheder uddybes herunder:

Alle aktive kontakter i en periode

Forskerservice anbefaler som udgangspunkt at afgrænse udtrækket, således at alle kontakter, der har været aktive i den ønskede periode, medtages. Dette kan opnås på følgende måder:

Specifik periode midt i LPRs historie:

Hvis du for eksempel er interesseret i data fra 1999-2005, anbefaler Forskerservice, at udtrækket afgrænses således:

indskrivningsdato <= 2005 og (udskrivningsdato >= 1999 eller udskrivningsdato er blank).

Ved denne afgrænsning opnår man, at alle igangværende kontakter i perioden kommer med, det vil sige også kontakter, som er startet inden periodens start eller er igangværende ved periodens slutning.

Specifik periode fra og med starten på LPRs historie:

Hvis du for eksempel er interesseret i data frem til og med 2015, anbefaler Forskerservice, at udtrækket afgrænses således:

indskrivningsdato <= 2015

På denne måde opnår du, at alle kontakter fra registerperiodens start kommer med, det vil sige også kontakter som er startet i årene før 1977. Ligeledes kommer der kontakter med, som er igangværende ved periodens slutning. Kontakter som er startet efter periodens slutning kommer ikke med.

Specifik periode til og med seneste opdatering:

Hvis du for eksempel er interesseret i data fra og med 1994, anbefaler Forskerservice, at udtrækket afgrænses således:

udskrivningsdato >= 1994 eller udskrivningsdato er blank

På denne måde udtrækkes alle igangværende kontakter i hele perioden inklusiv kontakter, der er startet inden periodens start eller er igangværende ved periodens slutning (seneste opdatering af LPR).

Alle kontakter indskrevet i en periode

Hvis udtrækket alene afgrænses på indskrivningsdato, vil udtrækket indeholde alle kontakter, der er påbegyndt i perioden.

Alle kontakter afsluttet i en periode

Udtrækket kan også afgrænses alene på udskrivningsdato, således at udtrækket indeholder alle kontakter, der er afsluttet i perioden. Ved denne afgrænsning er det vigtigt, at du angiver, at udtrækket desuden skal indeholde uafsluttede kontakter (kontakter uden en udskrivningsdato).

Individuel periodeafgrænsning

Det er desuden muligt at foretage en individuel periodeafgrænsning for hver person i en population. Det kunne for eksempel være data for en periode afgrænset med udgangspunkt i en indexdato (for eksempel en diagnosedato eller inklusionsdato).

Databrud

Du skal i forbindelse med tidsafgrænsningen være opmærksom på forskellige ændringer i LPR over tid. Det er ikke alle typer oplysninger, der dækker hele registerets periode, og undervejs er der skiftet klassifikationssystem for procedurer og diagnoser. De vigtigste databrud gælder:

- ICD8-systemet bruges for diagnoser i perioden 1977-1993. Fra 1994 og fremefter bruges det ICD-10 baserede SKS-system.
- Sundhedsstyrelsens operations- og behandlingsklassifikation for operationer (procedurer) bruges i perioden 1977-1995 og registreres i tabellen t_opr. Fra 1996 og fremefter bruges den danske udgave af NCSP (Nordic Classification of Surgical Procedures), og oplysningerne samles i tabellen t_sksopr.
- Undersøgelser og behandlinger registreres konsekvent fra og med 1999 og samles i tabellen t_sksube.

- › Ambulante kontakter og skadestuepatienter registreres fra og med 1994. Skadestuekontakter er ændret til akutte ambulante patientkontakter fra 2014.

Udtrækskriterier for kontakter

Du skal oplyse hvilke kontakter, der skal med i udtrækket. Følgende kriterier bruges typisk til at udvælge kontakter:

- › Alle kontakter for en kohorte/population
- › Alle kontakter, som har en given diagnose (dette kan enten være som aktionsdiagnose, bi-diagnose eller alle typer diagnoser).
- › Alle kontakter, som er registreret med en given kode for operation, undersøgelse eller behandling (procedurekoder).
- › Alle kontakter indskrevet på et givent sygehus eller afdeling.
- › Alle kontakter med en given bopælskommune eller –region.

Angivelse af diagnoser og procedurer

Når du angiver diagnoser og procedurer, kan du skrive dem helt ud, for eksempel: DK120A, DK120B, DK120C, DK120D.

Hvis du ønsker alle diagnoser under gruppen DK120, skriver du blot DK120*. Her angiver stjernen, at eventuelle underkoder medtages.

Du kan også angive intervaller: DK120A-DK120C (her vil DK120 og DK120D ikke være med).

Du skal angive både ICD10- og ICD8 diagnoser, hvis du skal have kontakter for hele perioden. Samme princip gælder for procedurer.

Uafsluttede ambulante kontakter

Nogle ambulante kontakter strækker sig over en længere periode. Det drejer sig blandt andet om forløb, hvor patienten løbende er til kontrol. Nogle ambulante kontakter bliver først afsluttet, når patienten dør. Skal disse med i udtrækket, skal du specifikt bede om uafsluttede kontakter.

Hovedtabel og satellittabeller

Oplysningerne i LPR er så mange, at registeret er opdelt i flere tabeller.

Hovedtabellen t_adm rummer alle administrative oplysninger. I tilknytning til t_adm findes en række satellittabeller, som hver især rummer detaljerede oplysninger om særlige dele af kontakten, for eksempel diagnoser (t_diag), operationer (t_opr og t_sksopr) eller undersøgelser og behandlinger (t_sksube).

Hovedtabel og satellittabeller knyttes sammen af en nøglevariabel - et unik recordnummer – sådan at alle rækker med et specifikt recordnummer i en given tabel rummer oplysninger om den unikke kontakt. K_reclnum er recordnummeret i t_adm, og v_reclnum er recordnummeret i satellittabellerne.

Recordnummeret er unikt i t_adm, det vil sige, at der findes én række pr. recordnummer, hvorimod det samme recordnummer kan forekomme i flere rækker i satellittabellerne (for eksempel hvis der på en kontakt er registreret mere end én diagnose).

Udtræk af individbaserede sundhedsdata fra LPR vil altid være bygget op på denne måde og leveret som adskilte tabeller, der kan kobles via recordnummeret.

Vigtigste tabeller og variable

Der findes en oversigt over tabeller og variable i dokumentationen. De vigtigste tabeller og variable nævnes dog her:

Tabeller

- t_adm: Administrative oplysninger: Personnummer, patienttype, indlæggelses- og udskrivningsdato og tidspunkt, sygehus- og afdelingskode med flere (1977 og fremefter)
- t_bes: Tabel med ambulante besøgsdatoer (1994 og fremefter)
- t_ulyk: Oplysninger om ulykke (2000 og fremefter)
- t_diag: Diagnoser (både ICD10 og ICD8 – 1977 og fremefter)
- t_opr: Operation efter den gamle Operations- og behandlingsklassifikation (1977-1995) – for kontakter, der er påbegyndt inden 1995, men afsluttet efter, er der også enkelte registreringer.
- t_sksopr: Operationer efter SKS operationsklassifikationen (1996 og fremefter) - K* koder i SKS systemet
- t_sksube: Undersøgelser og behandlinger efter SKS klassifikationen (1999 og fremefter).

Fra især den administrative tabel er det vigtigt at udvælge de konkrete variable, der er behov for i projektet, da den indeholder mange variable, hvor en del vil være irrelevante.

Variable

Her følger en oversigt over variable i de ovennævnte tabeller. Alle variable for de nævnte tabeller er med, men der findes flere tabeller i LPR end disse. For tabellen t_adm (administrative oplysninger) er for overblikkets skyld her angivet en gruppering, hvor man kan have brug for en

eller flere variable fra samme gruppe. For de øvrige tabeller er angivet grupperinger, hvor man typisk enten har brug for alle i gruppen eller ingen i gruppen.

Administrative oplysninger

Tabellen t_adm indeholder forskellige oplysninger om kontakten. Ud over disse informationer findes også tabellen t_bes med ambulante besøgsdatoer og t_ulyk med ulykkeskoder. Den sidstnævnte overtager fra 2000 ulykkeskoderne i t_adm tabellen.

Variablen k_recnum (i andre tabeller v_recnum) er nøgle for en kontakt. For hver kontakt vil følgende variable altid være unikke: CPR-nummer, patienttype, indlæggelsesdato, -time og -minut, sygehuskode og afdelingskode. Det betyder, at en person ikke kan være indlagt samme tidspunkt, samme sted og som samme patienttype med flere forskellige kontakter.

Tablet: t_adm

Gruppering	Variabel	Indhold	År
Nøglevariabel	k_recnum	Recordnummer (nøgle)	1977-
Personlige oplysninger	v_cpr	CPR nummer	1977-
	v_alddag	Alder i dage for 0-årige	1997-
	v_almdr	Alder i måneder	1997-
	v_alder	Alder	1977-
	c_sex	Køn	1977-
	c_kom	Bopælskommune	1977-
	c_amt	Bopælsamt	1994-
Basis kontaktoplysninger	c_pattype	Patienttype	1977-
Ind/ud tidspunkt	d_inddto	Indlæggelsesdato	1977-
	v_indtime	Indlæggelsestidspunkt	1977-
	v_indminut	Indlæggelsesminut	1994-
	v_udtime	Udskrivningstime	1994-
	d_uddto	Udskrivningsdato	1977-
Sted, basis	c_sgh	Sygehuskode	1977-
	c_afd	Afdelingskode	1977-
Sted, øvrige	c_spec	Speciale	1994-
	c_blok	Inddeling af speciale i blokke	1994-
	c_sghamt	Sygehusamt	1994-
Diagnose	c_adiag	Aktionsdiagnose	1977-
Henvisning	c_hsggh	Henvisende sygehus	2004-
	c_hafd	Henvisende afdeling	2004-
	d_hendto	Henvisningsdato	1977-
Ind fra / ud til	c_udtil	Udskrevet til	1977-1986
	c_indfra	Indlagt fra	1977-1986

Ulykkeskoder	c_ulykke	Ulykkeskode	1977-1986
	c_eakt	Ulykkeskode, aktivitet	1987-2003
	c_etra	Ulykkeskode, trafik	1987-2003
	c_ested	Ulykkeskode, sted	1987-2003
	c_emek	Ulykkeskode, mekanisme	1987-2003
	c_emodpart	Ulykkeskode, modpart	1995-2003
Ulykkeskoder	c_epart	Ulykkeskode, egenpart	1995-2003
Øvrige kontaktoplysninger	v_behdage	Behandlingsdage	1994-2001
	v_sengdage	Sengedage	1994-
	c_indm	Indlæggelsesmåde	1977-
	c_senstat	Senere statuskode	1977-1986
	c_andenbeh	Andenbehandling	1977-1986
	c_henm	Henvisningsmåde	1987-
	c_kontaars	Kontaktårsag	1987-
	c_udm	Udskrivningsmåde	1987-
	d_fusdto	Dato for forundersøgelse	2000-2003
	d_ebhdto	Data for endelig behandling	2000-2003

Tabel: t_bes (ambulante besøgsdatoer fra 1994-)

Gruppe	Variabel	Indhold
Grund	v_recnum	Recordnummer (nøgle)
	d_ambdto	Besøgsdato

Tabel: t_ulyk (ulykkeskoder fra 2000-)

Gruppe	Variabel	Indhold
Grund	v_recnum	Recordnummer (nøgle)
	c_art	Ulykkens art
	c_ulyk	Ulykkeskode
	c_tilulyk	Ulykkes tillægskode

Tabel: t_fritvalg (den udvidede fritvalgsordning fra 2004-)

Gruppe	Variabel	Indhold
Grund	v_recnum	Recordnummer (nøgle)
	c_fritvalg	Kode for fritvalgsordning

Diagnoser

Hvis der kun er behov for aktionsdiagnose, kan denne tages fra t_adm tabellen. Ellers er alle diagnoser (inklusive aktionsdiagnosen) i t_diag.

Tabel: t_diag (alle år)

Gruppe	Variabel	Indhold
Grund	v_recnum	Recordnummer (nøgle)
	c_diag	Diagnosekode (ICD8 / ICD10)
	c_diagart	Diagnoseart (aktions-, bidiagnose, tillægskode m.m.)
Med tillægskode	c_tildiag	ICD10 tillægskode (1995-)
	c_diagmod	ICD8 tillægskode (1977-1993) + ICD10 tillægskode 1994

Operationer

Tabel: t_opr (1977-)

Gruppe	Variabel	Indhold
Grund	v_recnum	Recordnummer (nøgle)
	c_opr	Operationskode
Med tillægskode	c_komb	Tillægskode (kombination)
Med stedsangivelse	c_osgh	Operationssygehus
	c_oafd	Operationsafdeling

Tabel: t_sksopr (1996 -)

Gruppe	Variabel	Indhold
Grund	v_recnum	Recordnummer (nøgle)
	c_opr	Operationskode
	c_oprart	Operationsart
Med tillægskode	c_tilopr	Tillægskode
Med stedsangivelse	c_osgh	Operationssygehus
	c_oafd	Operationsafdeling
Med tidsangivelse	d_odto	Operationsdato
	d_otime	Operationstime
	d_ominut	Operationsminut

Undersøgelser og behandlinger

Tabel: t_sksube (1999 -)

Gruppe	Variabel	Indhold
Grund	v_recnum	Recordnummer (nøgle)
	c_opr	Operationskode
	c_oprart	Operationsart
Med tillægskode	c_tilopr	Tillægskode
Med stedsangivelse	c_osgh	Operationssygehus
	c_oafd	Operationsafdeling
Med tidsangivelse	d_odto	Operationsdato
	d_otime	Operationstime
	d_ominut	Operationsminut

Betydning af prefix på variable

Alle variable er navngivet med et prefix. Det kan være k_, c_, v_ eller d_. Disse prefix angiver hvilket type indhold variabelen har.

Betydning:

k_: En nøglevariabel (*key*) – for eksempel k_recnum i t_adm.

c_: Indholdet er en kode (*code*) – for eksempel c_diag, der indeholder diagnosekoder.

v_: Indholdet kan aflæses som det pålydende (*value*) – for eksempel v_alder, der rummer alder i hele år.

d_: Indholdet er en dato (*date*) – for eksempel d_uddto, som er kontaktens afslutningsdato.