

Klinisk validering af LPR-baserede felter i DPCD-TOPICA - vurdering af omfang og betydning for datakvalitet

v / formand for DPCG og DPCD, overlæge Claus Frstrup i samarbejde med datamanager Brian Lind Olsen, kvalitetskonsulent Anne Zierau Kudsk Ragner og klinisk epidemiolog Henriette Engberg, DPCD Team i RKKP.

Endelig version per 09.10.24

Baggrund

Dansk Pancreas Cancer Database (DPCD) er baseret på DNKK-modellen, som genanvender allerede registrerede data i de danske nationale sundhedsregistre til etablering af patientforløb for patienter med kræft i bugspytkirtlen. Genanvendelse af registerdata indebærer, at rå-data fra LPR og Patologiregistret gennemgår en maskinel bearbejdning og opsætning i variable omkring forløb, patologisk grundlag for diagnose, kirurgisk og onkologisk behandling, via den såkaldte DNKK-algoritme.

De bearbejdede registerdata for hvert enkelt patientforløb præsenteres for klinikerne gennem TOPICA platformen ved log-in med unikt brugernavn og password. Flere af de centrale oplysninger omkring forløb og behandling kan rettes i TOPICA af klinikerne i forbindelse med gennemgang. Dette giver potentielt mulighed for mere valide data end ved en ren LPR-baseret database uden præsentation af data i en inddateringsplatform. Desuden kan der i TOPICA tilføjes manuelt registrerede data, som ikke aktuelt er mulige at indhente fra sundhedsregistre.

Det forventes, at DPCD skal overgå til at være ren LPR-baseret database uden TOPICA platformen, som led i en generel omlægning i RKKP. Dette kan potentielt medføre dårligere oversigt over patientforløb og behandling, samt færre rettelser og dermed potentielt ringere validitet af data.

Som forarbejde til omlægning er der derfor lavet en gennemgang af omfanget af kliniske rettelser og eksklusioner i DPCD-TOPICA databasen. Dette både for at anskueliggøre fremtidig validitet uden mulighed for direkte rettelser, og for at finde områder, hvor en kodevejledning til LPR kan mindske forekomst af fejlagtige data.

Opgørelsen er delt op i følgende områder: Forløb, Patologi, Onkologi og Kirurgi samt årsager til eksklusion fra databasen.

Opgørelsen omfatter perioden fra 2017 og frem til 1. september 2024, svarende til opstart af DNKK-modellen. Opgørelserne baseres på data til DPCD Årsrapport 2023/2024, som er udtrykt fra DPCD-TOPICA d. 6. september 2024.

Opgørelsen består i en sammenligning af data i variablene fra DNKK-modellen og de manuelt indtastede rettelser heraf i separate variable. Da databasen løbende opdateres med evt. rettelser i LPR, kan der ses kliniske rettelser, som efterfølgende automatisk er korrigeret i DNKK-modellens variable.

Data omkring *Eksklusioner fra databasen*

Databasen forsøger automatisk at frasortere patienter med forkert endelig diagnose både ud fra LPR-data og Patologiregistret. Ved den kliniske gennemgang er der mulighed for manuelt at ekskludere patienter fra databasen, hvis de ikke har pancreas-, papil- eller duodenalcancer. Antal eksklusioner har ligget mellem 10 og 20 patienter per år.

Der har været 135 (1,5%) manuelle eksklusioner i perioden 2017-2024. Den angivne årsag til eksklusion har været:

Forkert udgangspunkt af tumor	31
Forkert type af tumor (f.eks. NET)	64
Forkert kodning i LPR	33
Andet	6
Uoplyst	#

Konklusion

Den etablerede DNKK-algoritme er ikke fuldstændig nøjagtig. Men for ikke at risikere at udelukke reelle patienter er det svært at opnå 100% nøjagtighed.

Data omkring *Forløb*

Under *Forløb* er der mulighed for at korrigere oplysninger om *Udgangspunkt for tumor*, *Udredning i kræftpakke*, *Forløbsstartdato* og *Diagnosedato*.

Udgangspunkt for tumor:

Udgangspunkt for tumor er i perioden blevet korrigeret i 37 af 8415 forløb (<0,5%). Heraf er de 14 efterfølgende automatisk korrigeret via DNKK-modellen ud fra opdaterede data fra LPR. Aalborg Universitetshospital og Odense Universitetshospital har stået for næsten alle rettelser omkring *Udgangspunkt for tumor*.

Kræftpakkeforløb:

Det er alene Aalborg Universitetshospital, der har korrigeret oplysninger om kræftpakkeforløb. Der er lavet 142 (13%) rettelser i 1068 forløb. Heraf er 76 (>50%) efterfølgende automatisk korrigeret via DNKK-modellen ud fra opdaterede data fra LPR.

Forløbsstartdato og diagnosedato:

Det er alene Aalborg Universitetshospital, der har korrigeret datooplysninger. Forløbsstartdato er korrigeret i 172 tilfælde, hvorimod diagnosedato er korrigeret i 578 (54%) tilfælde af de 1068 forløb. Rettelsen af diagnosedatoen ligger indenfor +/- 7 dage hos 33% og +/- 14 dage hos 69%.

Konklusion

Samlet set er der yderst begrænsede rettelser omkring forløb, selv når der tages højde for at Aalborg Universitetshospital har stået for hovedparten af rettelserne.

Det vurderes ikke at overgang til ren LPR-database uden TOPICA vil give væsentlige problemer med forløbsdata.

Data omkring *Patologi*

Under *Patologi* er der mulighed for at korrigere *Grundlag for diagnose*, *Dato for patologi*, *Udgangspunkt*, *Materialetype* samt den endelige *histologiske konklusion*. Der har været få rettelser til alle punkter i 2017 og 2018, men siden da har der ikke været lavet kliniske rettelser.

Konklusion

Algoritmen til at fange og tolke data fra Patologiregistret er velfungerende og kan uden problemer overføres til en LPR-baseret tilgang uden TOPICA.

Data omkring Onkologi

Ved Onkologi har der været mulighed for korrektion af data omkring *Behandlingsregime* og *Dato for opstart af behandling*.

Derudover indeholder Onkologi-formularen manuelt indtastede data (variable) for *Behandlingsintention* og *Performancestatus*.

Behandlingsregime:

Der er registreret 6822 onkologiske formularer på 6037 patienter. Der er samlet set foretaget rettelser af behandlingsregime for 1274 (19%) formularer, dog er de 271 efterfølgende korrekt automatisk korrigeret via DNKK-algoritmen med opdaterede data fra LPR. Således er omfanget af reelle rettelser i alt 1003, svarende til 15%. Alle afdelinger foretager rettelser i væsentligt omfang, men ved Gødstrup Sygehus, Herlev Hospital og Odense Universitetshospital foretages rettelser i særlig høj grad (hhv. 30%, 27% og 19%).

Tabel 1 viser hvordan rettelserne fordeler sig. Det er især DNKK's bedømmelse af *gemcitabin+andet*, der rettes. Den hyppigste rettelse af data er fra *gemcitabin+andet* til *gemcitabin+nabpaclitaxel*. Dette er korrigeret for 392 onkologi formularer. *Gemcitabin monoterapi* er rettet til *gemcitabin+andet* for 167 formularer. *Best Supportive Care* er rettet til *andet* for 141 formularer. De øvrige rettelser er meget små i forhold til disse tre.

Tabel 1. Rettelser omkring *Behandlingsregime*. Rækker med DNKK værdi og kolonner med den kliniske rettelse / validering.

DNKK \ Klinisk Rettelse	1	2	3	4	5	6	7	I alt
1	1790	11	7	6	8	12	141	1975
2	4	1024	15	43	167	3	13	1269
3	0	3	319	#	0	#	5	330
4	3	5	0	538	21	3	3	573
5	7	43	#	392	530	#	3	979
6	#	#	#	#	3	1270	5	1283
7	#	#	10	11	3	38	348	413
I alt	1806	1089	354	994	732	1329	518	6822

1: *Best Supportive Care*, 2: *Gemcitabin*, 3: *Gemcitabin + Capacetabin*, 4: *Gemcitabin + Paclitaxel*, 5: *Gemcitabin + andet*, 6: *Folfirinox*, 7: *Andet*
#: $n=1$ eller $n=2$

Dato for opstart af behandling:

Datooplysninger rettes i meget begrænset omfang. Således er der kun 55 (<1%) rettelser ved de 6822 forløb, hvoraf 47 rettelser efterfølgende er korrekt automatisk korrigeret via DNKK-algoritmen med opdaterede data fra LPR.

Konklusion

Der foretages rettelser af onkologisk behandlingsregime i betydeligt omfang.

Overgang til ren LPR-baseret database uden TOPICA vil kræve en bedre kodevejledning og kodepraksis for LPR-data i de onkologiske afdelinger. Dette bør kombineres med optimering af DNKK-algoritmen til identifikation af behandlingsregime ud fra de observerede rettelser.

Endelig skal muligheden for indberetning af data til LPR for *Behandlingsintention* og *Performancestatus* undersøges.

Data omkring Kirurgi

For den kirurgiske behandling er der mulighed for at rette i *Type af behandling*, *Type af indgreb*, *Veneresektion*, *Arterieresektion*, *Re-operation* samt *Patologisk T- og N-stadium* samt afstande til resektionsflader.

Derudover registreres manuelt *Behandlingsintention*, *Årsager til ventetid*, *Type af indgreb ved palliation*, *M-stadium* og *Patologisk resektionsvurdering*.

Der er i perioden fra 2017 og frem registreret 2363 kirurgiske indgreb fordelt på 2195 patienter.

Type af behandling:

Der er meget få rettelser omkring *Type af behandling*. I alt 26 (1%) kirurgiske indgreb er rettet fra resektion til palliation, eksploration eller komplikation. Omvendt er 8 (<0,5%) kirurgiske indgreb rettet til resektion.

Type af indgreb:

Der er yderst få rettelser omkring den udførte resektion. Fordeling af indgreb fremgår af Tabel 2.

Tabel 2. Rettelser omkring *Type af resektion*. Rækker med DNKK værdi og kolonner med den kliniske rettelse / validering.

DNKK \ Rettelse	1	2	3	4	5	I alt
1	307	0	0	0	#	308
2	0	241	0	0	0	241
3	0	0	1237	0	0	1237
4	0	0	0	23	0	23
5	0	0	0	0	18	18
Missing	0	#	#	#	0	3
I alt	307	242	1238	24	19	1830

1: Distal, 2: Total, 3: Pankreatikoduodenektomi, 4: Lokalresektion, 5: Andet.

#: n= 1 eller n=2

Veneresektion:

Der har i perioden været 10 rettelser, hvor der er udført veneresektion uden at dette er fanget af DNKK-algoritmen. De 5 af rettelserne er dog automatisk fanget ved efterfølgende opdateringer af LPR-data. Samlet set er der lavet veneresektion ved 412 (23%) af de 1830 resektioner.

Arterieresektion:

Arterieresektion er først indført i 2019 som et DNKK-felt med indlæsning af data fra LPR. Der er derfor kun data fra 2019 og frem. Der er samlet set registreret 10 (1%) arterieresektioner ud af 1345 resektioner i perioden. De 5 er fanget af DNKK-algoritmen, mens der klinisk er tilføjet yderligere 5 arterieresektioner.

Re-operation:

Der er kun rettet i 7 tilfælde, hvoraf 4 er rettet til udførsel af re-operation, mens 3 er korrigeret til ingen re-operation.

Patologiske variable:

Der rettes i meget begrænset omfang i de patologiske variable. Efterfølgende opdateringer af data i DPCD-TOPICA viser desuden i meget stort omfang samme resultat som den kliniske rettelse.

Samlet konklusion

Den eksisterende DNKK-algoritme rammer for langt de fleste parametre meget præcist med ganske lille behov for kliniske rettelser. Den største forekomst af rettelser sker for den onkologiske behandling. Det er et område, hvor der på sigt forventes bedre overensstemmelse i forbindelse med overgang til Sygehusmedicinregisteret (SMR). Indtil dette bliver muligt, vil den underliggende DNKK-algoritme forventeligt kunne forbedres ud fra de observerede rettelser.

Ved overgang til en ren LPR-baseret database uden TOPICA platformen er der derudover en række manuelle variable, hvor data aktuelt ikke kan høstes fra LPR. Det skal derfor undersøges, om disse variable (data) vil kunne kodes og indberettes i LPR med mulighed for efterfølgende udtræk af data fra registeret.