

Use cases om brug af sundhedsdata til forskning og udvikling af innovative løsninger

Februar 2022



Danish
Life Science
Cluster

Indhold

Use-case 1: Adgang til behandlingsdata ved udvikling og kvalitetskontrol af algoritmer	3
Beskrivelse af generaliseret case	3
Overordnet problemstilling 1: Reguleringens styrende princip om adskillelse af forskning og behandling er en barriere for udvikling, afprøvning og anvendelse af algoritmer	4
Oplevet problemstilling 1A: Manglende adgang til at udvikle en algoritme baseret på data direkte fra behandlingssystemer	5
Oplevet problemstilling 1B: Kvalitetskontrol og tilpasning af algoritmen sker afkoblet fra behandlingssystemerne	5
Oplevet problemstilling 1C: Anvendelse af algoritmen i flere regioner forudsætter CE-mærkning	6
Use-case 2: Hypotesegenererende analyse af data fra bl.a. biologisk materiale	7
Beskrivelse af generaliseret case	7
Overordnet problemstilling 2: Reguleringens styrende principper om dataminimering og formålsbestemthed er en barriere for hypotese-genererende forskning i store datasæt, baseret på analyse af biologiske prøver	7
Oplevet problemstilling 2A: Fortolkningen af "ikke-omfattende" kortlægning er uklar og for restriktiv	8
Oplevet problemstilling 2B: Adgangen til at foretage analyser af biologisk materiale for kontrolpersoner er uklar og for restriktiv	8
Oplevet problemstilling 2C: Mangel på fleksible metoder til håndtering af information, transparens og medindflydelse til borgerne	9
Oplevet juridisk problemstilling 2D: Data fra "omfattende kortlægning" af biologisk materiale kan ikke genbruges	9
Oplevet juridisk problemstilling 2E: Fortolkning af krav om "væsentlig samfundsmæssig betydning" er uklar og for restriktiv.	9
Use-case 3: Mønstergenkendelse i journaloplysninger	11
Oplevet juridisk problemstilling 3A: Regionsrådets fortolkning af reguleringens krav om dataminimering og formålsbestemthed er en barriere for hypotese-genererende forskning baseret på journaldata	11
Oplevet juridisk problemstilling 3B: Regionsrådets fortolkning af krav om dataminimering og formålsbestemthed er uklare	12

Use-case 1: Adgang til behandlingsdata ved udvikling og kvalitetskontrol af algoritmer

Dette er en generaliseret beskrivelse af en anvendelse af sundhedsdata, der skal illustrere oplevede problemstillinger ved anvendelse af sundhedsdata til innovative og datadrevne analyser. Casen er baseret på forhold og oplevelser, som de er gengivet af forskere i interviews med PricewaterhouseCoopers Statsautoriseret Revisionspartnerselskab og Advokatfirmaet Poul Schmith (Kammeradvokaten). Enkelte forskere eller forskergrupper kan ikke tages til indtægt for de beskrivelser eller oplevede problemstillinger, som er gengivet i den generaliserede case.

Beskrivelse af generaliseret case

En forskergruppe, ledet af autoriserede sundhedspersoner, har udviklet en algoritme, der for en specifik sygdom understøtter klinikers mulighed for ved konsultation med god nøjagtighed at forudse en patients individualiserede risiko for komplikationer før behandling. Ved at anvende algoritmen er det dermed muligt at optimere patientbehandlingen ved at give mulighed for hurtig iværksættelse af behandling og tiltag for at reducere risikoen for alvorlige komplikationer. Algoritmen skal anvendes af autoriserede sundhedspersoner i forbindelse med aktuell behandling af patienter og selektion af patienter til inklusion i kliniske forsøg.

Der er en række væsentlige fordele for patienter, klinikere og sundhedsvæsenet herved.

Algoritmen er udviklet ved, for en gruppe patienter, at analysere historiske (retrospektive) data, der er indsamlet i forbindelse med patientbehandling. De pågældende data er registreret i patientjournaler og omfatter bl.a. diagnosekoder, patologikoder, laboratorietal og cellerelaterede data. Derudover anvendes der i udviklingen af algoritmen registerdata fra forskningsregistre, som er videregivet af Sundhedsdatastyrelsen, og som oprindeligt stammer fra journaldata. Der er gennem avanceret statistisk analyse identificeret et antal signifikante variable, der kan bidrage til at forudse en patients individualiserede risiko for komplikationer.

Forskningsprojektet til udvikling af algoritmen er godkendt af National Videnskabsetisk Komité efter komitéloven med dispensation fra kravet om informeret samtykke. Der er søgt om godkendelse af projektet efter komitéloven, fordi der i det samlede projekt også indgår en biobank til genetisk testning.

Når algoritmen skal anvendes til behandling, implementeres den i et EPJ-system. Det ønskes, at klinikerne i EPJ-systemet bliver præsenteret for algoritmens beregning for en patients risiko, baseret på den pågældende patients data i EPJ-systemet. Algoritmen er en matematisk formel for, hvordan bestemte variable (patientkarakteristika) har betydning for den individuelle patients risiko for komplikationer ved den givne sygdom. Algoritmen kan anvendes i forbindelse med aktuell behandling af en patient ved at indhente data om den konkrete patient fra EPJ i algoritmens formel, hvilket resulterer i en beregnet risiko for alvorlige komplikationer for patienten.

For at sætte algoritmen i produktion skal algoritmens datamodel efter udvikling manuelt mappes ("genidentificering af de enkelte data-variable") fra forskningsregistrenes data til EPJ-systemets data. Denne ændring er ressourcekrævende og giver risiko for fejl. Det er konstateret, at denne mapping er en kompleks opgave, givet algoritmens omfattende datamodel. Ønsket er, at algoritmen skal kunne implementeres i EPJ umiddelbart, baseret på de relevante forskningsdata den er bygget på – altså at forskningsdata og variable er en direkte 1:1-kopi af EPJ-data, så man undgår behovet for manuel mapping.

Når algoritmen efterfølgende skal kvalitetssikres og videreudvikles (trænes), er dette nødt til at foregå i et nyt forskningsprojekt, hvor historiske data om algoritmens vurderinger (prædiktioner) sammenholdes med prædiktioner fra de EPJ-integrerede prædiktioner (kvalitetssikring af reproducerbarhed).

Herudover er der behov for løbende opfølgning på om algoritmens prædiktioner fortsat er valide (monitorering af algoritmen efter implementering i EPJ-miljøet, altså opfølgning på om en ændring i den kliniske håndtering af sygdommen påvirker algoritmens prædiktion).

Når algoritmen efter kvalitetssikring (inklusive afprøvning i klinisk forsøg) og videreudvikling er ændret, skal denne implementeres/opdateres i EPJ-systemet på ny.

En yderligere datavariant, som foruden ovenstående giver udfordringer i brugen af sundhedsdata til blandt andet udvikling af algoritmer, er billedmateriale.

Algoritmen er som udgangspunkt kun til anvendelse i den pågældende region. Forskergruppens oplevelse er, at anvendelse i flere regioner forudsætter CE-mærkning af algoritmen eller individuel implementering i hver region (ifølge EUs Medical Device Regulation).

Overordnet problemstilling 1: Reguleringens styrende princip om adskillelse af forskning og behandling er en barriere for udvikling, afprøvning og anvendelse af algoritmer

I sundhedsreguleringen og databeskyttelsesreguleringen er der et styrende princip om en klar skelnen mellem forskning og behandling således, at behandlingsdata ikke umiddelbart kan anvendes til forskning. Omvendt må personoplysninger, der anvendes til statistiske og videnskabelige formål, som udgangspunkt ikke anvendes til andre formål, herunder til patientbehandling.

Denne klare skelnen mellem forskning og behandling vanskeliggør udviklingen og anvendelsen af algoritmer til beslutningsstøtte, da forskning og behandling "flyder sammen" i en algoritmes livscyklus med udvikling, anvendelse, evaluering og tilpasning. Samtidig er der af hensyn til behandlingskvalitet og patientsikkerhed behov for en høj kadence og "sømløshed" i den løbende evaluering og videreudvikling af algoritmen.

Konkret kommer denne overordnede problemstilling til udtryk i de nedenfor beskrevne oplevede problemstillinger.

Konsekvensen opleves at være, at udviklingen og anvendelsen af algoritmer til patientbehandling forsinkes, at algoritmer får en lavere kvalitet med ringere patientbehandling til følge, og at ressourceforbruget hos forskere til ikke-kliniske aktiviteter forøges.

Ønsket løsning: At reguleringen af anvendelsen af sundhedsdata tager højde for og understøtter, at forskningsudviklede algoritmer indgår i patientbehandling og forventeligt vil få en stadig større rolle her i de kommende år. Derudover ønskes der mulighed for juridisk rådgivning og fortolkning af diverse regler relateret til databehandling og forskning.

Oplevet problemstilling 1A: Manglende adgang til at udvikle en algoritme baseret på data direkte fra behandlingssystemer

Det opleves som en problemstilling, at algoritmer ikke kan udvikles ved at tilgå data direkte i behandlingssystemerne (prospektive data), men at det skal ske ved at tilgå historiske data (retrospektive data) fra forskningsregistre.

Det opleves, at dette introducerer en betydelig forsinkelse i adgangen til data i forbindelse med udviklingen af algoritmen – og en efterfølgende omkostning i form af tid, ressourcer og fejlkilder, når algoritmen skal mappes fra forskningsregistrenes data til behandlingssystemets data.

Ønsket løsning: At de relevante data fra behandlingssystemer kan tilgås direkte ved udvikling af algoritmer, eventuelt ved adgang til en altid opdateret kopi af de relevante data fra EPJ-systemet, naturligvis efter et projekt er godkendt.

Oplevet problemstilling 1B: Kvalitetskontrol og tilpasning af algoritmen sker afkoblet fra behandlingssystemerne

For at kunne anvende algoritmer i patientbehandlingen forsvarligt og effektivt er der behov for løbende og uden væsentlig forsinkelse at evaluere algoritmens korrekthed (reproducerbarhed mellem prædiktion på forskningsdata og på EPJ-data) og præcision (at prædiktionen ikke ændrer sig over tid, fx grundet ændringer i klinisk håndtering).

Der er behov for, at der i kvalitetskontrollen også anvendes data fra forskningsregistre. Som reguleringen forstås, må forskningsdata ikke videregives til kvalitetskontrol i behandlingssystemer. Derfor skal kvalitetskontrollen, med de nuværende reguleringer, ske i et forskningsmiljø (defineret som et forskningsprojekt og ikke et kvalitetssikringsprojekt).

Som reguleringen forstås, må data i behandlingssystemer alene bruges til patientbehandling, og kvalitetskontrol anses ikke for at være patientbehandling. Dermed skal kvalitetskontrollen ske i et efterfølgende forskningsprojekt baseret på historiske (retrospektive) data fra forskningsregistre, der sammenlignes med prædiktioner fra algoritmen i behandlingssystemet, som videregives til forskningsprojektet. Disse forskningsregisterbaserede data er typisk ét til halvandet år forsinkede, og dette medfører risiko for, at afvigelser i algoritmens korrekthed opdages sent – med risiko for alvorlige konsekvenser for behandlingskvaliteten.

Kvalitetskontrollen har bl.a. til formål at:

- sikre, at samme prædiktion af algoritmen er implementeret i behandlingssystemet og i forskningssystemet (reproducerbarhed i andet miljø)

- validere prospektivt, at algoritmen giver samme prædiktionsresultat som forventet i behandlingssystemet (fx opdage, hvis et analyseresultat, som indgår i algoritmen, gives i en anden enhed/et andet format) validere, at prædiktionsresultat svarer til outcome (detektere "drift" i outcome over tid, fx grundet ændring i klinisk håndtering).

Ønsket løsning: At relevante data fra behandlingssystemer kan tilgås direkte sammen med forskningsdata ved kvalitetskontrol af algoritmer med henblik på løbende kvalitetskontrol og tilpasning. Dette kan fx ske ved at forskergruppen kan tilgå en altid opdateret kopi af den relevante delmængde af data fra EPJ-systemet, som forskergruppen kan sammenstille med øvrige relevante data fra forskningsregistre. I en sådan model er det afgørende at afklare det juridiske og operationelle ansvar for, at kopiens data er korrekte i forhold til behandlingssystemets data. Dette kræver, at reguleringen tillader videregivelse af forsknings- og EPJ-data til kvalitetssikring af algoritmen.

Oplevet problemstilling 1C: Anvendelse af algoritmen i flere regioner forudsætter CE-mærkning

Algoritmen implementeres som udgangspunkt i én region i henhold til EU's Medical Device Regulation. Hvis algoritmen skal anvendes af flere regioner, forudsætter det en CE-mærkning af algoritmen. Årsagen er, at algoritmen i så fald klassificeres som medicinsk udstyr, jf. MDR-forordningen, hvor den kun må implementeres uden CE-mærkning i én region/enhed.

Dette opleves som en væsentlig ressourcemæssig og kompetencemæssig barriere for udbredelse af algoritmen og for dens bredest mulige validering.

Ønsket løsning: *[Denne problemstilling er uden for scope af analysen.]*

Use-case 2:

Hypotesegenererende analyse af data fra bl.a. biologisk materiale

Dette er en generaliseret beskrivelse af en anvendelse af sundhedsdata, der skal illustrere oplevede problemstillinger ved anvendelse af sundhedsdata til innovative og datadrevne analyser. Casen er baseret på forhold og oplevelser, som de er gengivet af forskere i interviews med PricewaterhouseCoopers Statsautoriseret Revisionspartnerselskab og Poul Schmith (Kammeradvokaten). Enkelte forskere eller forskergrupper kan ikke tages til indtægt for de beskrivelser eller oplevede problemstillinger, som er gengivet i den generaliserede case.

Beskrivelse af generaliseret case

En forskergruppe ønsker at gennemføre et forskningsprojekt med henblik på at identificere genetiske markører, der medfører øget risiko for en specifik sygdom hos en gruppe personer.

I analysen identificeres personer, der har været udsat for den specifikke sygdom, ved opslag i forskningsregistre, som forvaltes af en national myndighed.

For de identificerede personer, der har været udsat for den specifikke sygdom, foretages der genetisk analyse af biologiske prøver, som opbevares af en offentlig institution.

Tilsvarende foretages en genetisk analyse af biologiske prøver fra en sammenlignelig kontrolgruppe.

I analysen medtages en mindre del af de genetiske markører for at sikre, at anvendelse af genomdata klassificeres som "ikke-omfattende". Dette skyldes, at det af praktiske grunde ikke er muligt at indhente informeret samtykke fra de identificerede personer og kontrolgruppen – og ikke af videnskabelige grunde. Derfor søges der om godkendelse af forskningsprojektet efter bestemmelserne om "ikke-omfattende" kortlægning af genomet, som giver mulighed for dispensation fra kravet om informeret samtykke.

De genetiske data samkøres med datapunkter fra forskningsregistre.

Forskergruppen foretager herefter en statistisk analyse af, om der er signifikante forskelle i de undersøgte gener hos gruppen, der er udsat for den specifikke sygdom, i forhold til kontrolgruppen.

Overordnet problemstilling 2: Reguleringens styrende principper om dataminimering og formålsbestemthed er en barriere for hypotese-genererende forskning i store datasæt, baseret på analyse af biologiske prøver

I sundhedsreguleringen og databeskyttelsesreguleringen er der styrende princip og bestemmelser om dataminimering og formålsbestemthed, som har til formål at sikre, at der alene anvendes "nødvendige" data, og at de anvendes til klare og velbegrundede formål.

Dette opleves at være hindrende for udviklingen af nye behandlingsformer, der udnytter moderne datadrevne metoder til at analysere de omfattende datasæt, som bl.a. kommer fra genomsekventering og analyser af proteiner. Jo større populationer og jo flere parametre der kan indgå i analysen, desto større sandsynlighed er der for at identificere nye, signifikante og væsentlige sammenhænge, som kan medvirke til at udvikle nye behandlingsformer.

Samtidig kan det i denne type analyser være vanskeligt på forhånd at angive konkrete forventninger til analysens resultater, hvilke sammenhænge der forventes identificeret, og hvor stor en datamængde der vil være nødvendig.

Konkret kommer denne overordnede problemstilling til udtryk i de nedenfor beskrevne oplevede problemstillinger.

Ønsket løsning: At reguleringen af anvendelsen af sundhedsdata tager højde for og understøtter hypotese-genererende forskning, som udnytter de omfattende datamængder, der er til rådighed for medicinsk forskning og behandling i dag og i stigende omfang fremover.

Oplevet problemstilling 2A: Fortolkningen af "ikke-omfattende" kortlægning er uklar og for restriktiv

For at kunne få dispensation fra kravet om informeret samtykke, er det en forudsætning, at kortlægningen af den del af genomet, som projektet ønsker at analysere, klassificeres som "ikke-omfattende".

Det opleves, at der ikke er en klar og operationel definition af "ikke-omfattende" kortlægning af genomet hos de videnskabetiske komiteer. Dette giver usikkerhed i forhold til designet af analyser og flere iterationer med de godkendende myndigheder.

Samtidig opleves praksis hos de videnskabetiske komiteer som "for restriktiv", med den konsekvens at sandsynligheden for at finde signifikante sammenhænge i analyserne reduceres. Det opleves, at en restriktiv fortolkning følger et forsigtighedsprincip på grund af manglende anvisning fra politisk niveau af, hvordan loven ønskes implementeret.

Ønsket løsning: At de videnskabeligt set bedste genomdata kan tages i anvendelse, og at brugen af genetiske data fastlægges fra politisk niveau på en klar og mindre restriktiv måde.

Oplevet problemstilling 2B: Adgangen til at foretage analyser af biologisk materiale for kontrolpersoner er uklar og for restriktiv

For at kunne gennemføre analyser er der ofte behov for at inkludere en sammenlignelig kontrolgruppe, som ikke har været eksponeret for den pågældende helbredsmæssige hændelse.

Det opleves som vanskeligt at opnå godkendelse hertil, når der for kontrolgruppen også skal foretages analyser af biologisk materiale.

Ønsket løsning: At adgangen til at inkludere kontrolpersoner og analyser af deres biologiske materiale defineres klarere og mindre restriktivt, fx gennem genbrug af data fra andre analyser.

Oplevet problemstilling 2C: Mangel på fleksible metoder til håndtering af information, transparens og medindflydelse til borgerne

Hvis en kortlægning af biologisk materiale klassificeres som "omfattende", eller hvis der ikke kan gives dispensation, er der i reguleringen¹ krav om informeret samtykke fra de personer, hvis biologiske materiale skal indgå i analysen.

Det opleves som uforholdsmæssigt vanskeligt og ressourcekrævende at indhente informeret samtykke fra en større gruppe personer. Samtidig opleves det, at der kommer en betydelig social slagside i, hvilke personer der giver informeret samtykke, hvilket har konsekvenser for anvendeligheden af analysens resultater.

Kravet om informeret samtykke opleves endvidere at være i kontrast til befolkningens generelle tilslutning til, at deres prøver og analyser anvendes til forskning i bedre behandlingsformer.

Ønsket løsning: At der etableres en juridisk og eventuelt teknisk løsning til mere smidig håndtering af samtykke, fx adgang for borgere til at give et generelt samtykke eller mulighed for en "opt-out"-model for samtykke og/eller øget transparens, i forhold til hvad data fra biologiske prøver anvendes til.

Oplevet juridisk problemstilling 2D: Data fra "omfattende kortlægning" af biologisk materiale kan ikke genbruges

Det følger af lovgivningen, at analyser af biologisk materiale alene må bruges til det formål, der oprindeligt er givet godkendelse til. Det kan fx være en hel eller delvis gensekventering af en vævsprøve.

Det opleves, at der derfor ikke kan foretages supplerende analyser på datagrundlag fra allerede, gennemførte studier. Denne problemstilling opleves, såfremt der er tale om projekter, som er godkendte af National Videnskabsetisk Komité, og som indeholder en "omfattende kortlægning af genomet". De projekter, som derimod er godkendte af de regionale videnskabsetiske komitéer, og som omfatter en "ikke-omfattende kortlægning af genomet", kræver derimod ikke en særskilt tilladelse til genbrug af biologisk materiale.

Det følger deraf, at det enkelte projekt skal foretage "sin egen" analyse af biologisk materiale, hvilket er en tids- og ressourcemæssig omkostning, da den nye genetiske analyse blot vil frembringe de samme data igen. Dette reducerer samtidig mængden af de biologiske prøver.

Ønsket løsning: At der etableres en juridisk og eventuelt teknisk løsning for, at data fra analyser af biologisk materiale kan genbruges til andre væsentlige og relevante formål.

Oplevet juridisk problemstilling 2E: Fortolkning af krav om "væsentlig samfundsmæssig betydning" er uklar og for restriktiv.

Det er en forudsætning for at kunne lave statistiske analyser på biologisk materiale, at National Videnskabsetisk Komité vurderer, at forskningsprojektet har "væsentlig samfundsmæssig interesse".

¹ Komitéloven

Det opleves, at definitionen af "væsentlig samfundsmæssig interesse" er uklar, og at National Videnskabsetisk Komité i nogle tilfælde fortolker dette for restriktivt i et etisk perspektiv. Det opleves desuden, at det kan virke vilkårligt, hvornår en given sygdom/et givet sundhedsspørgsmål vurderes at være tilstrækkeligt væsentligt for samfundets interesse. Dette gør sig især gældende for projekter, som er grundvidenskabelige.

Ønsket løsning: At "væsentlig samfundsmæssig betydning" defineres klarere, ensartet og mindre restriktivt.

Use-case 3: Mønstergenkendelse i journaloplysninger

En forskergruppe ønsker at udvikle en algoritme, der kan anvendes af sundhedspersoner som et beslutningsstøtteværktøj i den løbende patientbehandling.

Algoritmen skal støtte sundhedspersonerne i at identificere markører i en bestemt type prøvesvar, som sundhedspersonerne skal være særligt opmærksomme på og hurtigt tage stilling til. Algoritmen skal identificere markører, der er af særlig interesse på baggrund af patienters aktuelle behandlinger og sygdomshistorik. Dette kan være i form af værdier i prøvesvaret, som er uden for normalområdet, eller værdier, der kræver særlig opmærksomhed, grundet patientens nuværende og tidligere diagnoser.

Oprindeligt ønskede forskningsgruppen at udvikle algoritmen ved at foretage analyser af et større antal patientjournaler for at gøre algoritmen i stand til at finde sammenhænge mellem sygdomskategorier og markører gennem avanceret mønstergenkendelse. Datagrundlaget vil være diagnoser og prøvesvar, der er registreret i patientjournalerne.

Regionsrådet, som er godkendende myndighed, afviste ansøgningen herom, med henvisning til at datasættet var for omfattende, og at formålet ikke var tilstrækkeligt specifikt. Forskeren tilpassede derfor ansøgningen således, at antallet af patientjournaler og antallet af variable i analysen var væsentligt mindre, og så der alene indgik få specifikke sygdomskategorier i analysen.

Forskningsgruppen har nu opnået godkendelse af det tilpassede projekt i det pågældende regionsråd.

Oplevet konsekvens

Forskeren vurderer, at konsekvensen af det reducerede datasæt og det reducerede antal sygdomskategorier, der analyseres sammenhænge for, er en væsentligt reduceret potentiel værdi for sundhedspersoner og patienter af forskningen og den algoritme, der søges udviklet.

Oplevet juridisk problemstilling 3A: Regionsrådets fortolkning af reguleringens krav om dataminimering og formålsbestemthed er en barriere for hypotesegenererende forskning baseret på journaldata

Det opleves, at regionsrådets fortolkning af reguleringens krav om dataminimering og formålsbestemthed – og dermed den godkendende myndigheds krav om indskrænkning af datagrundlag og analyserede sygdomskategorier for at godkendte projektet – væsentligt forringer kvaliteten og anvendeligheden af et forskningsprojekt, der har til formål at give klinikere bedre beslutningsstøtte – og afledt heraf bedre patientbehandling.

Ønsket løsning: At reguleringen og den godkendende myndigheds fortolkning heraf understøtter, at der kan foretages hypotesegenererende analyser i omfattende datasæt.

Oplevet juridisk problemstilling 3B: Regionsrådets fortolkning af krav om dataminimering og formålsbestemthed er uklare

Det opleves, at regionrådets fortolkning af reguleringens krav om dataminimering og formålsbestemthed er uklar og at det endeligt godkendte omfang af data og inkluderede diagnoser i analysen i et vist omfang er vilkårligt. Udbredte folkesygdomme har stor diversitet og kræver især en stor mængde patientmateriale.

Ønsket løsning: At der er klarere vejledning således, at forskere bedre kan tilrettelægge projekter efter lovgivningens rammer og så ansøgningsforløbet med den godkendende myndighed bliver mere effektiv.