

Vejledning for validering af data fra kliniske kvalitetsdatabaser

Juni 2017

Projektdeltagere

Projektet udarbejdes af RKKP's daværende Kompetencecenter for Epidemiologi og Biostatistik (KCEB)-Nord, Klinisk Epidemiologisk Afdeling, Olof Palmes Allé 43-45, 8200 Aarhus N.

Projektgruppen består af:

Alma Becic Pedersen, overlæge, dr. med., ph.d., klinisk lektor, E-mail: abp@clin.au.dk

Søren Paaske Johnsen, overlæge, ph.d., klinisk lektor, E-mail: spj@clin.au.dk

Projektet er finansieret efter ansøgning om projektmidler fra Regionernes Udviklingspulje for Klinisk Kvalitet.

Indeværende rapport er udarbejdet af projektgruppen – med undtagelse af afsnit om hjemmel, der er skrevet af RKKP.

Indhold

Resume og anbefalinger	4
Baggrund.....	5
Formål.....	5
Datavaliditet – definition af de vigtigste begreber	6
Metoder for validitet undersøgelser	8
1. Dækningsgrad af patientregistreringen	8
Eksempel 1.1. Dækningsgrad af patientregistrering i forhold til et uafhængigt register	8
Eksempel 1.2. Dækningsgrad af patientregistrering i forhold til flere uafhængige registre.....	9
Eksempel 1.3. Dækningsgrad af patientregistrering baseret på flere uafhængige registre	10
Eksempel 1.4. Dækningsgrad af patientregistrering for databaser med primær indsamling via administrative registre.....	11
2. Datakomplethed af variabler / indikatorer i databasen	11
Eksempel 2.1. Datakomplethed af variabler	11
Eksempel 2.2. Datakomplethed af kvalitetsindikatorer.....	13
3. Datakvalitet	15
Eksempel 3.1. PPV og NPV undersøgelse i forhold til patientjournal.....	15
Eksempel 3.2. PPV og NPV undersøgelse i forhold til algoritme	15
Eksempel 3.3. Er patientjournalen en god guldstandard for PPV og NPV undersøgelse?.....	16
Stikprøvestørrelse	17
Hjemmel	18
Reference.....	18

Resume og anbefalinger

- Kendskab til datavaliditeten er en forudsætning for en meningsfuld anvendelse og fortolkning af data fra de kliniske kvalitetsdatabaser.
- Datavaliditeten af kliniske kvalitetsdatabaser omfatter tre aspekter: 1) dækningsgrad af patient registrering (har vi alle relevante patienter i databasen?), 2) datakomplethed af variabler/kvalitetsindikatorer (har vi registreret alle relevante variabler og kvalitetsindikatorer for hver enkelt patienter?), og 3) datakvalitet (afspejler de registrerede data virkeligheden?).
- Forskellige metoder kan bruges til at validere data fra kliniske databaser, afhængig af hvilket aspekt af datavaliditeten skal det undersøges.
- Et valideringsprojekt udføres ifølge en protokol, hvor formålet med projektet beskrives detaljeret og klart (hvilket aspekt af datavaliditeten man vil undersøge), design, metoden, dataindsamlingsprocessen, datakilder, stikprøvens størrelse og udtagelse, statistiske analyser, samt projektets begrænsninger.
- Et valideringsprojekt vil typisk ikke kunne give svar på alle de forskellige aspekter af datavaliditeten på en gang.
- Datavaliditet ændrer sig over tiden. Derfor er validerings af kliniske kvalitetsdatabaser en kontinuerlig proces og valideringsprojekter i en database bør gennemføres regelmæssigt, eksempelvis hvert 3.-5. år, i lighed med opdatering af indikatorsættet.
- Valideringsprojekt kan med fordel indgå som led i større projekter, f.eks. et ph.d. projekt, hvor en forsker kan få bedre kendskab til indsamling af data, kan bidrage til udvikling af data og kan bidrage til videnskabelig udnyttelse og fortolkning af data.
- Ved journal gennemgang, afhænger valget af stikprøvestørrelse af den minimale forskel af to målinger man ønsker at detektere, præcision af estimerer man ønsker at opnå, praktiske forhold og gennemførlighed. Disse forhold skal vejes imod udgifterne.
- Baggrunden for stikprøvestørrelse behøver ikke nødvendigvis at være en formel statistisk analyse, men snarere en pragmatisk afvejning mellem at opnå en rimelig statistisk præcision af de estimerede sensitivitet, specificitet og positive og negative prædiktive værdier, og de ressourcer der er til rådighed. Det afgørende er at ambitionerne med valideringsarbejdet er afstemt med de ressourcer der afsættes til opgaven.

- RKKP kan med fordel lægge loft over budgettet for valideringsprojekter, og understøtte at medfinansiering kan indgå i projekterne.

Baggrund

En klinisk kvalitetsdatabase kan defineres som: "Et register, der indeholder udvalgte kvantificerbare indikatorer, som kan belyse dele af eller den samlede kvalitet af sundhedsvæsnets indsats og resultater for en afgrænset patientgruppe med udgangspunkt i det enkelte patientforløb" (1). Det er ca. 60 kliniske databaser som modtager bevilling fra Regionerne. Disse databaser omfatter næsten alle specialer, og inkluderer både sygdoms- og indsatsdatabaser. Flertallet af databaser er etableret indenfor de sidste 10 år og de anvender forskellige datakilder, som Klinisk Måle System (KMS), Landspatientregisteret, EPJ systemer osv.

Formålet med de kliniske kvalitetsdatabaser er at forbedre kvaliteten af forebyggelse, diagnostik, behandling og rehabilitering, udgøre dokumentation for den kliniske og politiske ledelse, bidrage til gennemsigtighed i sundhedsvæsenet via information til borgere og patienter, samt udgøre en del af den nationale forskningsinfrastruktur. Ud over kvalitetsindikatorer, indeholder en klinisk kvalitetsdatabase som regel en række prognostiske faktorer (f.eks. body mass index, boligforhold, alkohol og rygning, komorbiditet), samt data vedr. den underliggende sygdom (f.eks. sværhedsgrad) og behandling (f.eks. type af operation, anæstesitype, implantat type, operationstid), som kan bruges til databasens formål.

Dårlig datavaliditet udgør en alvorlig trussel mod brug af data fra kliniske databaser og bør tages alvorligt. Kendskab til datavaliditeten er en forudsætning for en meningsfuld anvendelse af de indsamlede data.

Formål

Formålet med denne vejledning er at beskrive de vigtigste aspekter af datavaliditet og de metoder som kan bruges til at undersøge validiteten af data fra kliniske kvalitetsdatabaser.

Datavaliditet – definition af de vigtigste begreber

Til at beskrive datavaliditeten anvendes en række kliniske epidemiologiske mål inklusiv sensitivitet, specificitet, positiv og negativ prædiktiv værdi (2). I database sammenhæng, anvender man ofte betegelse sensitivitet/specificitet, når man taler om dækningsgrad af registrering eller datakomplethed og PPV/NPV, når man taler om datakvaliteten.

Som guld standard kan anvendes en anden klinisk kvalitetsdatabase, et administrativt register, eller patientjournaler, alene eller i kombination. Guld standarden bruges til at sammenligne de observerede data med data, som anses for at være tæt på sandheden eller i det mindste den bedste tilgængelige data. Guld standarden kan dog have begrænsninger, som vi vil beskrive lidt nemmere længere ned i teksten.

Table 1. Sammenligning af data i databasen med "guld standard".

		"Guld standard"	
		Sygdom til stede (+)	Sygdom ikke til stede (-)
Database	Sygdom til stede (+)	Sand positiv (a)	Falsk positiv (b)
	Sygdom ikke til stede (-)	Falsk negativ (c)	Sand negativ (d)

Sensitiviteten forstås som andelen af syge personer, der er blevet registreret i den pågældende database datasæt.

$$\text{Sensitivitet} = a/a+c$$

Specificiteten forstås som andelen af personer uden den relevante sygdom, der er registreret som ikke havende sygdommen i den pågældende database datasæt.

$$\text{Specificiteten} = d/b+d$$

Sensitiviteten og specificiteten anvendes i database sammenhæng når man taler om dækningsgrad af registrering (er alle relevante patienter registreret?) eller datakomplethed (er alle relevante data for hver patient registreret?).

Dækningsgraden af registrering siger noget om, i hvilket omfang det er lykkedes at få inkluderet alle relevante patienter i datasættet uden samtidig at få inkluderet andre patienter, som måtte have været indlagt eller ambulantly behandlet i samme periode med

andre sygdomme. Med andre ord, databasens dækningsgrad er en procentvis angivelse af antallet af registrerede patienter i databasen ud af det samlede antal patienter, som opfylder inklusionskriterierne. Dækningsgrad opgøres på populationsniveau som en del af afrapporteringen fra de kliniske kvalitetsdatabaser. Dækningsgraden af registrering bør iht. BEK nr. 975 af 28/06/2016 fra Sundhedsdatastyrelsen og basiskrav for kliniske kvalitetsdatabaser være over 90% (3), dvs. over 90% af patienter, som er behandlet for pågældende sygdom skal være registreret i database datasæt. Dette er vigtig for at vurdere klinisk kvalitet af behandling af disse patienter og for at sikre, at resultaterne er repræsentative.

Datakompletheden er lige så vigtig som dækningsgraden af registrering og kan opgøres på både variabel og indikator niveau. Datakomplethed af en variabel er en procentvis angivelse af antallet af registrerede patienter i databasen som har fået registreret denne variabel, ud af det samlede antal patienter, som variabelen er relevant for. Datakompletheden af indikatorer kan angives som antallet af registrerede patienter i databasen med tilstrækkelige informationer til at indgå i en indikatoropgørelse ud af det samlede antal patienter, som er relevante i opgørelsen. Det er ikke specifikke krav vedr. datakompletheden, men høj komplethed vil sikre mere valid fortolkning af resultaterne

Positive prædiktiv værdi (PPV) forstås som den andel af de syge personer, som er reelt syge.

$$PPV = a/a+b$$

Negative prædiktiv værdi (NPV) forstås som den andel af de ikke syge personer, som er registreret i database datasæt som ikke syge.

$$NPV = d/c+d$$

PPV og NPV betyder i database sammenhæng, i hvilket omfang de syge eller ikke syge som er registreret i database datasæt er reelt syge eller ikke syge. Endvidere, det er vigtigt at vide om de variabler som er registreret for hver patient er korrekte, dvs. det er vigtigt at vide for eksempel om patient er opereret på grund af hofte fraktur, som registreret i databasen eller om patienten i virkelighed var opereret på grund af en anden hofte diagnose. Det betyder med andre ord nøjagtigheden af indberettede variabler, eller variabel overensstemmelse med virkeligheden.

Metoder for validitet undersøgelser

1. Dækningsgrad af patientregistreringen

Dækningsgraden af patientregistrering i en klinisk kvalitetsdatabase kan undersøges på forskellige måder, ved at sammenligne indberetninger fra en klinisk kvalitetsdatabase med indberetninger af de samme patienter fra et andet uafhængigt register, eller fra en kombination af flere uafhængige registre. I praksis kan man anvende forskellige uafhængige datakilder, så som Landspatientregister (4), Patologiregisteret, eller Cancerregisteret (5, 6).

Eksempel 1.1. Dækningsgrad af patientregistrering i forhold til et uafhængigt register

I et validering studie blev kompletheden af registreringen af total hoftealloplastik operationer i Dansk Hoftealloplastik Register (DHR) undersøgt (7). DHR er en landsdækkende klinisk kvalitetsdatabase etableret i 1995, som indeholder oplysninger om patienter som har fået primær eller/og revision total hoftealloplastik operation. Guldstandard i valideringsstudiet var Landspatientregisteret. Dækningsgraden af patient registreringen var 94,1%, idet 26129 operationer var registreret i DHR og 27757 operationer var registreret i DHR og/eller Landspatientregister. Dækningsgraden var lige høj for mænd og kvinder, og på tværs af aldersgrupper. Dækningsgraden var dog kun 86,7% for lav volumen sygehuse sammenlignet med 95,8% for høj volumen sygehuse, og 93,9% for primær operationer sammenlignet med 81,4% for revision operationer. Desuden, viste studiet at dækningsgraden var stigende over tiden.

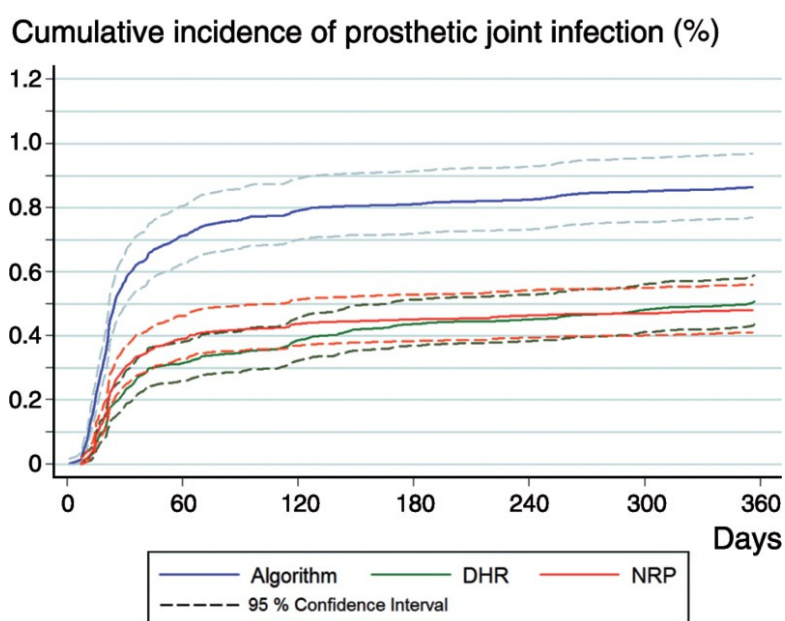
Studiets resultater tyder på, at det er vigtigt at opgøre dækningsgrad på forskellige niveau, for forskellige variable og med regelmæssige intervaller, idet den ikke var konstant over tid. Studiet viste endvidere at ca. 700 patienter var registreret i DHR men ikke Landspatientregisteret, hvilket understreger at Landspatientregister er ikke perfekt guld standard.

Direkte sammenligning af en database med et andet uafhængigt register giver ikke mulighed for at beregne specificitet.

Eksempel 1.2. Dækningsgrad af patientregistrering i forhold til flere uafhængige registre

Hvis det ikke er muligt at sikre sig at alle patienter bliver identificeret via ét uafhængigt register, kan man kombinere flere uafhængige registre. Det er vigtigt at udarbejde en præcis algoritme som beskriver, hvordan data skal kombineres. En algoritme er en kombination af variabler som giver bedre mulighed for at identificere patienter uden direkte kontakt med patienten. Som variabler kan man bruge diagnosekoder, operationskoder, patologikoder, biokemiske koder og andre register data.

Figur 1. Incidensen af revisioner pga. dyb infektion identificeret via DHR (Dansk Hoftalloplastik Register), NRP (Landspatientregister) og algoritme.



Et valideringsstudie undersøgte dækningsgraden af registrering af revisioner pga. dyb infektion i DHR (8). Selvom dækningsgraden af revisioner generelt var høj i et tidligere studie, var det uklart i hvilket omfang de specifikke årsager til revision, især dyb infektion var komplet registreret. Mistanken var opstået pga. af den måde information vedr. dyb infektion bliver registreret i DHR, nemlig før et mikrobiologisk svar er tilgængelig for den læge som skal indberette revisionsårsagen. Forskerne udviklede en algoritme baseret på data vedr. mikrobiologi (9), indløste recepter (10), og klinisk biokemiske data (11), og anvendte denne algoritme som "guld standard" for at finde patienter med revision pga. dyb infektion. Studiet har vist at dækningsgraden af registreringen af revisioner pga. dyb infektion i DHR var cirka 66%, dvs. 34% af de reelle revisioner pga. dyb infektion ikke kunne

identificeres i DHR. Dækningsgraden steg til 90% ved at kombinere DHR med mikrobiologi databasen alene.

Eksempel 1.3. Dækningsgrad af patientregistrering baseret på flere uafhængige registre

Tendensen i dansk sundhedsvæsenet i forhold til kliniske kvalitetsdatabaser har været at undgå dobbeltregistreringer. Et stigende antal databaser benytter sig således af data indberettet til allerede eksisterende administrative registre. Det er i den forbindelse nødvendigt at validere den algoritme, som det påtænkes at den specifikke database skal anvende til at identificere patienter.

Et eksempel fra Dansk Colorektal Cancer Group database kan illustrere udfordringen.

Formålet med studiet var at udvikle og validere en algoritme for identifikation af patienter med kolorektal cancer recidiv baseret på danske administrative registre (12). Algoritmen var udviklet ved at kombinere metastase og kemoterapi koder fra Landspatientregisteret med koder fra Patologiregisteret. Guld standard var 355 danske patienter med kolorektal cancer diagnosen registreret i Dansk Colorektal Cancer Group database i perioden 2001-2011. Disse patienter var fulgt op igennem 2012 for recidiv diagnosen. Algoritmen havde sensitivitet på 95% (60 ud af 63 patienter med recidiv blev identificeret) og specificitet på 97% (282 ud af 290). Disse data viser at ved at anvende algoritmen ville databasen være i stand til at identificere 95% af patienterne med recidiv og kun 3% af patienter vil være falsk positive. Studiet viste at en kombination af danske register data kan bruges til at identificere cancer recidiv patienter.

Tablet 2. Sammenligning af kolorektal cancer recidiv algoritme med "guld standard"

Algoritme	"Guld standard"	
	Recidiv (+)	Recidiv (-)
Recidiv (+)	60 (a)	10 (b)
Recidiv (-)	3 (c)	282 (d)

Eksempel 1.4. Dækningsgrad af patientregistrering for databaser med primær indsamling via administrative registre

Et stigende antal databaser, f.eks. Dansk Register for Kronisk Obstruktiv Lungesygdom, Databasen for Akutte Hospitalskontakter, Dansk Intensiv Database og Dansk Tværfagligt Register for Hoftenære Lårbensbrud bygger på data indhentet fra Landspatientregisteret og der er ikke altid andre velegnede oplagte uafhængige datakilder til undersøgelse af dækningsgraden af registrering. Fraværet af en registrering i disse databaser tolkes derfor typisk som fravær af en given tilstand eller procedure. Der er dog behov for løbende validering mod data fra de enkelte sygehuse, herunder fra kliniske logistiksystemer som Patient Administrative System (PAS) eller Ledelses Information systemer (LIS), velvidende at disse datakilder heller ikke er perfekte.

Screenings databaser, som f.eks. Dansk Kvalitetsdatabase for Livmoderhalskræftscreening (DKLS), baserer sig på data fra Patologidatabanken (Patobanken), der modtager data fra alle landets patologiafdelinger og privatpraktiserende speciallæger i patologi, herunder data vedrørende celleprøver fra livmoderhalsen. Oplysningerne i Patobanken opdateres dagligt ved overførsel af data fra både patologiafdelingernes og de privatpraktiserende patologers patologiesystemer. Det betyder, at alle celleprøver fra livmoderhalsen løbende registreres i Patobanken. Det giver en registreringsprocent tæt på 100% i Patobanken, samt en dækningsgrad i DKLS databasen på næsten 100 %, idet der kan være manglende registreringer på grund af fejl i indberetningerne.

2. Datakomplethed af variabler / indikatorer i databasen

Det er ikke nok at patienten er indberettet til databasen, men også at de variabler som burde være indberettet for hver enkelt patient er indberettet.

Eksempel 2.1. Datakomplethed af variabler

Hvis databasen anvender et selvstændigt indtastningssystem som eksempelvis Klinisk Måle System (KMS), er der mulighed for at opgøre komplethed for hver enkelt variabel ved at tabulere antal af patienter som har fået registreret en specifik variabel over antal af patienter som burde have fået registrering. Disse simple opgørelser bør foretages

regelmæssigt og mindst én gang om året i forbindelse med udarbejdelse af årsrapporten. I tabel 3 nedenunder præsenteres en sådan opgørelse fra DHR, som viser at 7,7% af patienter registreret med primær operation i 2015 har manglende registrering af data vedr. varighed af tromboseprofylakse. Det ses endvidere at andelen af patienter med manglende registrering af varighed af tromboseprofylakse er faldende siden 2010. Ligeledes viser den efterfølgende tabel 4, at 0,1% af patienterne registreret i DHR i perioden 1995-2015 manglede oplysning om operativ adgang ved primær operation.

Tabel 3. Kompletthed af registrering af variabel ”varighed af tromboseprofylakse” i DHR.

Tromboseprofylakse [dage] for primær operationer

	2010		2011		2012		2013		2014		2015	
	n	%	n	%	n	%	n	%	n	%	n	%
Antal dage												
0-5	1963	21.8	2210	24.8	2630	29.7	2928	32.4	4034	42.8	6267	64.8
11-20	287	3.2	265	3.0	159	1.8	159	1.8	93	1.0	56	0.6
6-10	2875	32.0	2753	30.9	2648	29.9	3190	35.3	3007	31.9	2426	25.1
>20	2335	26.0	3017	33.9	2772	31.3	2039	22.6	1520	16.1	183	1.9
Missing	1527	17.0	657	7.4	651	7.3	726	8.0	761	8.1	742	7.7
I alt	8987	100.0	8902	100.0	8860	100.0	9042	100.0	9415	100.0	9674	100.0

Tabel 4. Kompletthed af registrering af variabel ”operativ adgang” i DHR.

Operativ adgang ved primæroperationer

Operativ adgang	1995-2013		2014		2015		Total	
	n	%	n	%	n	%	n	%
Bagre	118081	90.8	9045	96.1	9303	96.2	136429	91.5
Lateral	8571	6.6	355	3.8	361	3.7	9287	6.2
Minimal invasiv								
surgery	2688	2.1	0	0	0	0	2688	1.8
Forreste	408	0.3	10	0.1	6	0.1	424	0.3
Andet	174	0.1	0	0	0	0	174	0.1
Missing	143	0.1	5	0.1	4	0.0	152	0.1
I alt	130065	100.0	9415	100.0	9674	100.0	149154	100.0

Hvis databasen er baseret på administrative registre som f.eks. Landspatientregisteret, kan der stadig undersøges komplementhed af variabler ved at bruge de specifikke SKS diagnose-, procedure- og tillægskoder. I tabel 5 nedenunder har man opgjort komplementhed af variabelen "boligforhold" i Dansk Tværfagligt Register for Hoftenære Lårbensbrud. Opgørelsen viser at 194 (17%) af patienterne i Region Sjælland har ikke fået registreret SKS-kode for boligforhold (missing eller ikke udfyldt), og at 29 (3%) af patienterne har ikke oplyst deres boligforhold.

Tabel 5. Komplementhed af variabel "boligforhold" i Dansk Tværfagligt Register for Hoftenære Lårbensbrud.

Prognostisk faktor - Boligforhold	SKS-kode ikke indberettet		Egen Bolig (ZRSB01)		Egen bolig i tilknytning til institution (ZRSB01A)		Døgninstitution (ZRSB04)		Ingen oplysninger om boligforhold (ZRSB09)		I alt N
	Antal	%	Antal	%	Antal	%	Antal	%	Antal	%	
<i>Sjælland</i>	194	17	725	64	44	4	147	13	29	3	1139
<i>Roskilde og Køge Sygehuse ortopedkirurgisk afdeling</i>	56	17	187	57	12	4	54	16	21	6	330
<i>3800J40 - HOL O-kir. Afd.</i>	13	5	202	82	7	3	23	9	0	0	245
<i>3800S30 - SLA O-kir. Afd.</i>	16	5	222	73	18	6	49	16	0	0	305
<i>3800X00+V40 - NFS, O-kir.</i>	109	42	114	44	7	3	21	8	8	3	259

Eksempel 2.2. Datakomplethed af kvalitetsindikatorer

Datakomplethed af en kvalitetsindikator kan opgøres, hvor "ikke oplyst" medfører eksklusion fra en indikatoropgørelse. For databaser, hvor mangelfuld registrering ("ikke oplyst") sidestilles med "ikke opfyldt", vil datakomplethed ikke meningsfuldt kunne opgøres. I Dansk Apopleksiregister opgøres eksempelvis indikatoren "Andel af patienter med akut apopleksi, der får vurderet ernæringsrisiko senest 2. indlæggelsesdag." Ernæringsrisiko kan registreres i følgende kategorier: Angives som "Ja", "Nej", "Ikke relevant" eller "Uoplyst" samt dato (dd-mm-åå). I tabel 6 nedenunder kan man vurdere komplementheden af denne indikator, idet 349 (3%) af patienterne har mangelfuld registrering ("ikke oplyst") og er dermed ekskluderet fra indikatoropgørelsen.

Table 6. Kompletthed af indikator "Andel af patienter med akut apopleksi, der får vurderet ernæringsrisiko senest 2. indlæggelsesdag." i Dansk Apopleksiregister.

	<i>Standard opfyldt</i>	<i>Tæller/nævner</i>	<i>Uoplyst antal (%)</i>	<i>Aktuelle år</i> 2015 % 95% CI	<i>Tidligere år</i> 2014 % (95% CI)	<i>Tidligere år</i> 2013 % (95% CI)
Danmark	nej	8931 / 10376	349 (3)	86 (85 - 87)	87 (86-88)	86 (85-86)
Hovedstaden	nej	2597 / 3114	107 (3)	83 (82 - 85)	86 (85-87)	88 (86-89)
Sjælland	ja	1587 / 1733	47 (3)	92 (90 - 93)	91 (89-92)	89 (87-90)
Syddanmark	nej	1756 / 1972	123 (6)	89 (88 - 90)	87 (85-88)	89 (88-91)
Midtjylland	nej	2068 / 2499	10 (0)	83 (81 - 84)	86 (85-87)	81 (79-83)
Nordjylland	nej	923 / 1058	62 (6)	87 (85 - 89)	88 (86-90)	75 (72-78)

Enkelte databaser, som Dansk Tværfagligt Register for Høftenære Lårbensbrud, har ikke skelnet mellem "ikke oplyst" og "ikke udfyldt" i indikator opgørelserne. Afdelinger med mange "ikke udfyldte" indikatorer er således tvunget til at foretage indberetning til databasen for at deres resultaterne ikke fremstår som dårlige.

Content validity eller på dansk indholdsvaliditet er et andet aspekt vedr.

kvalitetsindikatorerne, som det kan være relevant at undersøge. Content validity afspejler hvorvidt kvalitetsindikatorer anvendt i databasen måler alle de relevante aspekter af behandlings kvaliteten, som de skal måle. Med andre ord, kan de udvalgte kvalitetsindikatorer bruges til at måle behandlingskvaliteten og identificere udvikling over tiden. Dette kunne undersøges ved at spørge en international ekspert gruppe som består af eksperter fra diverse faggrupper, involveret i behandling af de specifikke patienter. Alternativt, kan der foretages sammenligning med kvalitetsindikatorer som anvendes i lignende databaser internationalt. For eksempel, har en undersøgelse af Riksstroke, som er det national svenske apopleksiregister vist at 18 ud af 22 kvalitetsindikatorer blev identificeret af internationale eksperter som værende relevante for måling af kvaliteten af behandlingen af apopleksi patienter. En anden undersøgelse viste at 14 ud af 15 undersøgte kvalitetsindikatorer er anvendt i mindst 2 europæiske apopleksi registre, hvilket afspejler at der synes at være international konsensus om, hvilke indikatorer der kan anvendes til at monitorere kvaliteten af apopleksibehandlingen (13).

3. Datakvalitet

PPV og NPV angiver sandsynlighed for, at en patient hhv. har eller ikke har den relevante diagnose, operation eller har fået den behandling, såfremt vedkommende er registreret i databasen. PPV og NPR undersøges i forhold til en guldstandard, eksempelvis patientjournalen eller en valideret algoritme. Der er foretaget talrige valideringsstudier af danske kvalitetsdatabaser af denne karakter igennem årene.

Datakvalitet undersøgelser kan give svar ikke kun om omfanget af eventuel misklassifikation af for eksempel diagnoser, men også om misklassifikationen er systematisk, dvs. om misklassifikationen afhænger af et bestemt patient karakteristisk som alder, køn eller kalender tidspunkt. Dette er især vigtig hvis data skal bruges til analytiske studier, dvs. studier som analyserer årsag til forekomst af sygdom eller prognosen af sygdom.

Eksempel 3.1. PPV og NPV undersøgelse i forhold til patientjournal

I et valideringsstudie af Dansk Apopleksiregister blev PPV og NPV af registrering af medicinske komplikationer blandt patienter indlagt med akut apopleksi i perioden fra 2003 til 2006 undersøgt i forhold til patientjournaler. En tilfældig prøve af 589 patienter (7,3% af hele populationen) blev udtaget og et registreringsskema med enstydige definitioner af de medicinske komplikationer blev udviklet og brugt systematisk ved journal gennemgangen. PPV af alle medicinske komplikationer tilsammen registreret i Dansk Apopleksiregister var 71,7% (95% CI: 67,4%-75,8%) med variation i PPV fra 39,0% (95% CI: 24,2%-55,5%) for tryksår til 87,1% (95% CI: 78,0%-93,5%) for pneumoni. NPV af alle medicinske komplikationer tilsammen var 91,3% (95% CI: 90,0%-92,2%) med variation fra 71,6% (95% CI: 67,2%-75,7%) for pneumoni til 98,9% (95% CI: 97,5%-99,6%) for lungeemboli (14).

Eksempel 3.2. PPV og NPV undersøgelse i forhold til algoritme

I et andet valideringsstudie af DHR, blev PPV og NPR af registrering af revision på grund af dyb infektion undersøgt i forhold til en tidligere udviklet og valideret algoritme, baseret på andre registre og journal data (15). PPV og NPV blev opgjort i forhold til køn, operationsår, tiden til revision, type af bakterier, CRP niveau, og pusdannelse (Tabel 7). Studiet viser at

estimerterne er lave, men var de samme i alle grupper, d.v.s. at misklassifikationen ikke var differentiel.

Tabel 7. Datakvalitet af revisioner pga. dyb infektion i DHR.

Table I. Sensitivity, specificity, positive predictive value and negative predictive value with 95% confidence interval for the diagnosis of 'periprosthetic joint infection' (PJI) in the Danish Hip Arthroplasty Register (DHR)

	No. of revisions n	Reported as PJI in DHR n	Defined as PJI by algorithm n	Sensitivity % (range)	Specificity % (range)	Positive predictive value % (range)	Negative predictive Value % (range)
Overall	1382	232	267	67.0 (61.0 to 72.6)	95.2 (93.8 to 96.4)	77.2 (71.2 to 82.4)	92.3 (90.7 to 93.8)
Time to revision after primary THA							
Early < 3 mths	544	127	159	64.8 (56.8 to 72.2)	93.8 (90.9 to 96.0)	81.1 (73.2 to 87.5)	86.6 (82.9 to 89.7)
Delayed 3 to 24	441	78	76	75.0 (63.7 to 84.2)	94.2 (91.3 to 96.4)	73.1 (61.8 to 82.5)	94.8 (91.9 to 96.8)
Late > 24 mths	397	27	32	59.4 (40.6 to 76.3)	97.8 (95.7 to 99)	70.4 (49.8 to 86.2)	96.5 (94.1 to 98.1)
Years of reporting							
2005 to 2006	162	27	37	56.8 (39.5 to 72.9)	95.2 (89.8 to 98.2)	77.8 (57.7 to 91.4)	88.1 (81.5 to 93.1)
2007 to 2008	297	65	61	80.3 (68.2 to 89.4)	93.2 (89.2 to 96.1)	75.4 (63.1 to 85.2)	94.8 (91.1 to 97.3)
2009 to 2010	412	70	88	64.8 (53.9 to 74.7)	96.0 (93.2 to 97.8)	81.4 (70.3 to 89.7)	90.9 (87.4 to 93.8)
2011 to 2012	511	70	81	64.2 (52.8 to 74.6)	95.8 (93.5 to 97.5)	74.3 (62.4 to 84.0)	93.4 (90.7 to 95.6)
Gender							
Male	634	116	121	72.7 (63.9 to 80.4)	94.5 (92.2 to 96.3)	75.9 (67.0 to 83.3)	93.6 (91.2 to 95.6)
Female	748	116	146	62.3 (53.9 to 70.2)	95.8 (93.9 to 97.3)	78.4 (69.9 to 85.5)	91.3 (88.8 to 93.4)
Type of bacteria							
Virulent group	118	92	111	82.9 (74.6 to 89.4)	100 (59.0 to 100)	100 (96.1 to 100)	26.9 (11.6 to 47.8)
Opportunistic	173	68	134	50.7 (42.0 to 59.5)	100 (91.0 to 100)	100 (94.7 to 100)	37.1 (27.9 to 47.1)
Spore forming	5	2	4	50.0 (6.8 to 93.2)	100 (2.5 to 100)	100 (15.8 to 100)	33.3 (0.8 to 90.6)
C-reactive protein							
Elevated	569	191	218	70.6 (64.1 to 76.6)	89.5 (85.8 to 92.5)	80.6 (74.3 to 86.0)	83.1 (78.9 to 86.7)
Normal	412	18	18	33.3 (13.3 to 59.0)	97.0 (94.8 to 98.4)	33.3 (13.3 to 59.0)	97.0 (94.8 to 98.4)
Presence of purulence							
Presence of purulence	47	43	34	94.1 (80.3 to 99.3)	15.4 (1.9 to 45.4)	74.4 (58.8 to 86.5)	50.0 (6.8 to 93.2)
Absence of purulence	456	59	29	62.1 (42.3 to 79.3)	90.4 (87.2 to 93.0)	30.5 (19.2 to 43.9)	97.2 (95.1 to 98.6)

Eksempel 3.3. Er patientjournalen en god guldstandard for PPV og NPV undersøgelse?

Mange validering studier har anvendt patientjournalen som guldstandard, hvor alle processer omkring behandlinger af patienter bør være dokumenteret. Desværre har enkelte valideringsstudier påvist at anvendelse af patientjournal som guldstandard kan være problematisk. Pedersen et al (2012) validerede data i Den Nationale Skizofrenidatabase (16). I studiet blev der gennemgået i alt 196 journaler (1% af alle patienter registreret i perioden 2004-2009). Studiet viste bl.a. at patientjournaler i varierende grad manglede dokumentation for centrale oplysninger om patient behandlingen , fra 1% for antipsykotisk medicin, 41% for selvmord risiko vurdering, til 54% for psyko-uddannelse. Studiet konkluderede at dokumentation af behandlingen af skizofreni patienter på danske psykiatriske afdelinger i patientjournaler ikke er tilfredsstillende og advarer imod anvendelsen af patientjournaler som guldstandard for valideringsprojekter.

Stikprøvestørrelse

Uanset om det drejer sig om valideringsstudier eller andre typer af studier, er det en god ide at tænke på og beregne stikprøvestørrelse. Ofte kan vi af praktiske eller ressourcemæssige årsager ikke inkludere alle patienter med bestemt sygdom i en undersøgelse og vi er nødt til at udvælge en tilfældig gruppe af patienter, som vi vil undersøge. Stikprøvestørrelsen er vigtig i alle typer af undersøgelser, hvor vi gerne vil udtale os om hele populationen baseret på en tilfældigt udvalgt gruppe af patienter.

For at bestemme den mest hensigtsmæssige stikprøvestørrelse kan med fordel anvendes specifikke programmer (17). Flere sådanne programmer er gratis tilgængelige eksempelvis Ken Rothman's Episheet program (<http://www.epidemiolog.net/studymat/>).

De hidtil publicerede valideringsstudier fra danske kliniske kvalitetsdatabaser har anvendt ganske forskellige stikprøvestørrelser. I et valideringsstudie fra Dansk Korsbånd Register (DKKR), sygehus journaler på stikprøve af 5% (i alt 250 journaler) af patienterne i DKRR i en 7 års periode gennemgået (18). I Dansk Apopleksiregister blev der gennemgået journaler på 7% (i alt 589 journaler) af patienterne inkluderet i databasen i gennem en 4 årig periode (14), mens der i Dansk Akut Leukæmi database blev gennemgået i alt 260 journaler, svarende til 10% af patienterne i en 1 års periode (19).

Sammenhæng mellem stikprøvestørrelse og præcision er illustreret i nedenstående tabel 8, med en hypotetisk undersøgelse med fire prøver varierende fra 100 til 1000 personer, fra en population på 1000 personer og med forskellige PPV værdier (d.v.s. 80, 85 og 90%).

Tabel 8. Sammenhæng mellem stikprøvestørrelse og præcision af PPV værdier.

Stikprøvestørrelse	PPV (95% CI)	PPV (95%CI)	PPV (95% CI)
100	85 (78,36 – 91,64)	80 (72,56 - 87,44)	90 (84,42 – 95,58)
200	85 (80,57 – 89,43)	80 (75,04 – 84,96)	90 (86,28 – 93,72)
500	85 (82,79 – 87,21)	80 (77,52 – 82,48)	90 (88,14 – 91,86)
800	85 (83,89 – 86,11)	80 (78,76 – 81,24)	90 (89,07 – 90,93)

Stikprøvestørrelsen afhænger af den minimale forskel af to målinger vi gerne vil være i stand til at detektere, præcisionen af de estimater vi ønsker at opnå, praktiske forhold og gennemførlighed. Udgifterne til data indsamling på en stikprøve skal endvidere tages i betragtning. Desuden er det vigtigt at være opmærksom på to forhold: 1) en valideringsundersøgelse af en database vil ikke kunne give svar på alle de forskellige aspekter af datavaliditeten som er blevet præsenteret og diskuteret i de tidligere afsnit af denne rapport; 2) at datavaliditet ofte ændrer sig over tid. Derfor er validering af kliniske kvalitetsdatabaser en kontinuerlig proces og valideringsprojekter i en database bør gennemføres hver 3.-5. år, i lighed med opdatering af indikatorsættet.

Hjemmel

Validering af databasernes indhold kan ske inden for rammerne af godkendelser og anmeldelser af databaserne, hvis det sker ved godkendte databehandlere og/eller behandlingsansvarlige.

Hvis der som led i validering trækkes på data, der ikke er godkendt indhold i databaserne, skal det ske via et godkendt forskningsprojekt. Læs mere her:

<http://sundhedsdatastyrelsen.dk/da/forskerservice>

Reference

1. Kjaergaard J. Kvalitetsudvikling i Sundhedsvæsenet: Munksgaard; 2001.
2. Rothman KJ. Epidemiology. An Introduction.: Oxford University Press; 2002 2002.
3. Regioner D. Basiskrav for landsdækkende kliniske kvalitetsdatabaser2007.
4. Schmidt M, Schmidt SA, Sandegaard JL, Ehrenstein V, Pedersen L, Sorensen HT. The Danish National Patient Registry: a review of content, data quality, and research potential. Clin Epidemiol. 2015;7:449-90.
5. Storm HH. Completeness of cancer registration in Denmark 1943-1966 and efficacy of record linkage procedures. Int J Epidemiol. 1988;17(1):44-9.
6. Storm HH, Michelsen EV, Clemmensen IH, Pihl J. The Danish Cancer Registry--history, content, quality and use. Dan Med Bull. 1997;44(5):535-9.

7. Pedersen A, Johnsen S, Overgaard S, Soballe K, Sorensen H, Lucht U. Registration in the Danish Hip Arthroplasty Registry: Completeness of total hip arthroplasties and positive predictive value of registered diagnosis and postoperative complications. *Acta Orthop Scand*. 2004;75(4):434-41.
8. Gundtoft PH, Overgaard S, Schonheyder HC, Moller JK, Kjoergaard J, Pedersen CB. The "true" incidence of prosthetic joint infection after 32,896 primary total hip arthroplasty: A prospective cohort study. *Acta Orthop*. 2014.
9. Schonheyder HC, Sogaard M. Existing data sources for clinical epidemiology: The North Denmark Bacteremia Research Database. *Clin Epidemiol*. 2010;2:171-8.
10. Johannesdottir SA, Horvath-Puho E, Ehrenstein V, Schmidt M, Pedersen L, Sorensen HT. Existing data sources for clinical epidemiology: The Danish National Database of Reimbursed Prescriptions. *Clin Epidemiol*. 2012;4:303-13.
11. Grann AF, Erichsen R, Nielsen AG, Froslev T, Thomsen RW. Existing data sources for clinical epidemiology: The clinical laboratory information system (LABKA) research database at Aarhus University, Denmark. *Clin Epidemiol*. 2011;3:133-8.
12. Lash TL, Riis AH, Ostefeld EB, Erichsen R, Vyberg M, Thorlacius-Ussing O. A validated algorithm to ascertain colorectal cancer recurrence using registry resources in Denmark. *Int J Cancer*. 2015;136(9):2210-5.
13. Soderman P. On the validity of the results from the Swedish National Total Hip Arthroplasty register. *Acta Orthop Scand Suppl*. 2000;71(296):1-33.
14. Ingeman A, Andersen G, Hundborg HH, Johnsen SP. Medical complications in patients with stroke: data validity in a stroke registry and a hospital discharge registry. *Clin Epidemiol*. 2010;2:5-13.
15. Gundtoft PH, Pedersen AB, Schonheyder HC, Overgaard S. Validation of the diagnosis 'prosthetic joint infection' in the Danish Hip Arthroplasty Register. *Bone Joint J*. 2016;98-B(3):320-5.
16. Pedersen CG, Gradus JL, Johnsen SP, Mainz J. Challenges in validating quality of care data in a schizophrenia registry: experience from the Danish National Indicator Project. *Clin Epidemiol*. 2012;4:201-7.
17. Flahault A, Cadilhac M, Thomas G. Sample size calculation should be performed for design accuracy in diagnostic test studies. *J Clin Epidemiol*. 2005;58(8):859-62.
18. Rahr-Wagner L, Thillemann TM, Lind MC, Pedersen AB. Validation of 14,500 operated knees registered in the Danish Knee Ligament Reconstruction Register: registration completeness and validity of key variables. *Clin Epidemiol*. 2013;5:219-28.
19. Ostgard LS, Norgaard JM, Severinsen MT, Sengelov H, Friis L, Jensen MK, et al. Data quality in the Danish National Acute Leukemia Registry: a hematological data resource. *Clin Epidemiol*. 2013;5:335-44.