

Formaliserte konsensusprosesser, veiledning ved vanskelige valg?

Notat

Kunnskapssenteret

Juli 2009

 kunnskapssenteret

Indikatorer er en kilde til kunnskap om pasientsikkerhet. Hvilke indikatorer man velger å benytte avgjør hvilket bilde som blir tegnet. Hvordan valg av indikatorer foregår og blir dokumentert er derfor viktig informasjon når man skal forstå og tolke det bilde som indikatorene gir. • Flere land og internasjonale organisasjoner benytter en eller annen form for formaliserte konsensusprosesser i utvikling og valg av kvalitetsindikatorer for helsetjenesten. I dette notatet ser vi nærmere på prinsippene bak konsensusprosesser og beskriver en amerikansk metode og to europeiske verktøy som er blitt brukt til dette formålet. • I Norge er det liten tradisjon for å benytte formaliserte konsensusprosesser for å dokumentere valg av indikatorer. Ved å bygge på de erfaringer som andre land og organisasjoner har gjort på dette feltet kan man ved relativt enkle midler ta dette verktøyet i bruk. Både ved utvikling av medisinske kvalitetstetsregister og nasjonale indikatorer kan konsensusprosesser benyttes. Det vil i første omgang medføre en del ekstra kostnader, men på sikt vil kunnskapsgrunnlaget bedres og en vil spare tid både ved utvikling *(fortsetter på baksiden)*

Nasjonalt kunnskapssenter for helsetjenesten
Postboks 7004, St. Olavs plass
N-0130 Oslo
(+47) 23 25 50 00
www.kunnskapssenteret.no
Notat. ISBN 978-82-8121-281-7

Juli 2009

kunnskapssenteret

(fortsettelsen fra forsiden)

og validering av indikatorer. • Dokumentasjonen av konsensusprosesser gir fyldig informasjon om den enkelte indikator og hvordan den blir vurdert av et fagmiljø. Styrker og svakheter ved indikatoren blir dokumentert og er viktig bakgrunnsinformasjon i fortolkning av indikatorer når den blir benyttet. Kostnader ved datainnsamling kan føre til at en indikator ikke velges, og er viktig informasjon når man skal forstå helheten av informasjon om et indikatorsett. Åpenhet om valg og konsekvensene av dem er et viktig ledd i å bedre forståelsen og nytten av kvalitetsindikatorer for helsetjenesten.

Tittel	Formaliserte konsensusprosesser, veiledning ved vanskelige valg?
Institusjon	Nasjonalt kunnskapssenter for helsetjenesten
Ansvarlig	Geir Bukholm, <i>avdelingsdirektør</i>
Forfattere	Saunes IS, <i>forsker</i> Deilkås E, Kristoffersen M, Lyngstad M, og Nyen B
ISBN	978-82-8121-281-7
Rapport	Notat
Prosjektnummer	912
Rapporttype	Notat
Antall sider	39
Oppdragsgiver	Nasjonal enhet for pasientsikkerhet
Sitering	

Nasjonalt kunnskapssenter for helsetjenesten fremskaffer og formidler kunnskap om effekt av metoder, virkemidler og tiltak og om kvalitet innen alle deler av helsetjenesten. Målet er å bidra til gode beslutninger slik at brukerne får best mulig helsetjenester. Senteret er formelt et forvaltningsorgan under Sosial- og helsedirektoratet, uten myndighetsfunksjoner. Kunnskapssenteret kan ikke instrueres i faglige spørsmål.

Nasjonalt kunnskapssenter for helsetjenesten
Oslo, juni 2006

Oppsummering

Indikatorer er en kilde til kunnskap om pasientsikkerhet. Hvilke indikatorer man velger å benytte avgjør hvilket bilde som blir tegnet. Hvordan valg av indikatorer foregår og blir dokumentert er derfor viktig informasjon når man skal forstå og tolke det bilde som indikatorene gir.

Flere land og internasjonale organisasjoner benytter en eller annen form for formaliserte konsensusprosesser i utvikling og valg av kvalitetsindikatorer for helsetjenesten. I dette notatet ser vi nærmere på prinsippene bak konsensusprosesser og beskriver en amerikansk metode og to europeiske verktøy som er blitt brukt til dette formålet.

I Norge er det liten tradisjon for å benytte formaliserte konsensusprosesser for å dokumentere valg av indikatorer. Ved å bygge på de erfaringer som andre land og organisasjoner har gjort på dette feltet kan man ved relativt enkle midler ta dette verktøyet i bruk. Både ved utvikling av medisinske kvalitetsregister og nasjonale indikatorer kan konsensusprosesser benyttes. Det vil i første omgang medføre en del ekstra kostnader, men på sikt vil kunnskapsgrunnlaget bedres og en vil spare tid både ved utvikling og validering av indikatorer.

Dokumentasjonen av konsensusprosesser gir fylldig informasjon om den enkelte indikator og hvordan den blir vurdert av et fagmiljø. Styrker og svakheter ved indikatoren blir dokumentert og er viktig bakgrunnsinformasjon i fortolkning av indikatorer når den blir benyttet. Kostnader ved datainnsamling kan føre til at en indikator ikke velges, og er viktig informasjon når man skal forstå helheten av informasjon om et indikatorsett. Åpenhet om valg og konsekvensene av dem er et viktig ledd i å bedre forståelsen og nytten av kvalitetsindikatorer for helsetjenesten.

Notatet er et resultat av arbeidsgruppens arbeid, og anbefalingene er uttrykk for synspunktene til arbeidsgruppens medlemmer.

Sammendrag

Kunnskap om pasientsikkerhet kommer fra mange kilder. Pasientsikkerhetsindikatorer kan gi systematisk kunnskap om enkelte aspekt ved pasientsikkerhet. Når indikatorer skal velges for å måle pasientsikkerhet er det flere forhold som skal belyses: Hvordan velges indikatorene? Hvem velger hvilke indikatorer som skal benyttes? Hvilket krav til dokumentasjon skal det være i prosessene som fører fram til valget?

I dette notatet beskriver vi tre ulike varianter av konsensusprosesser og drøfter hva som kan være egnet for å velge indikatorer for pasientsikkerhet i Norge.

Standardisert konsensusprosess

Formaliserte konsensusprosesser kan kort beskrives som en tretrinns evalueringssprosess. Kjennetegn ved standardiserte konsensusprosesser er den systematiske tilnærmingen med klare krav til dokumentasjon av kunnskapsgrunnet, og hvilke kriterier som brukes i evalueringen og dokumentering av både prosess og resultat.

Først trinn i konsensusprosessen er å identifisere og framskaffe dokumentasjon for indikatoren. Det gjøres som regel av et fagmiljø med kompetanse på systematisk oppsummering og er en kritisk vurdering og sammenfatning av tilgjengelig litteratur. Konsensusprosesser inkluderer også studier som ikke tilfredsstillende de samme strenge krav til metode som systematiske kunnskapsoppsummeringer.

Et ekspertpanel vurderer dokumentasjonen med hensyn til validitet og antatt nytteverdi av indikatorene. Validitet og reliabilitet vurderes, i lag med tilgang på data. En eventuell rapporteringsbyrde må sees i sammenheng med nytteverdien. Denne vurderingen er andre trinn i prosessen.

Sammensetning av ekspertpanelet er viktig. En kan velge eksperter bare innenfor et medisinsk og helsefaglig område, men de fleste panel er bredere sammensatt med kliniske spesialister og faglige ressurspersoner fra flere fagområder.

Ekspertpanelet møtes sjelden ved den første evaluering av indikatorene. Paneldokumentene sendes til deltakerne med en forklaring av videre saksgang. Dokumentene skal inneholde litteraturgjennomgang, liste over indikatorer, definisjoner og in-

struksjoner. Vurderingen av kunnskapsgrunnlaget gjøres ved en spørreskjema - undersøkelse hvor indikatorene rangeres med hensyn til validitet som kvalitetsmål og mulighetene for å samle inn de data som er spesifisert i indikatoren.

Det tredje og siste trinnet i prosessen er å samle deltakerne i panelet for å gjennomføre en endelig evaluering. Først samles panelet, og resultatene av spørreskjema - undersøkelsen presenteres. Etter plenumsdiskusjoner rangerer ekspertpanelet tiltaket/indikatorene på nytt. Formålet med metodikken er ikke å framtvinge konsensus, men å dokumentere hva panelet er enig eller uenig om når det gjelder vurdering av indikatorene. Indikatorer som oppnår en gitt grad av oppslutning velges og indikatorer som det er klar uenighet om utelukkes.

En amerikansk metode og to europeiske verktøy

I denne rapporten gis en innføring i en amerikansk metode for konsensusprosess. Det er Rand/UCLA sin metodikk som blant annet er mye benyttet av Agency for Healthcare Research and Quality og den ligger også til grunn for gjennomføring av konsensusprosesser i National Quality Forum (NQF) i USA.

Deretter beskrives to europeiske verktøy, Stepwise Assessment Indicator Framework Approach (SAIFA) som har vært benyttet i et pasientsikkerhetsprosjekt i EU, og Appraisal of Indicators through Research and Evaluation (AIRE) som er utviklet i Nederland. De to siste verktøyene er valgt ettersom de er laget spesifikt for indikatorvurdering.

Bruksområde

Innenfor helsetjenesteforskning og administrasjon er formaliserte konsensusprosesser benyttet når det foreligger manglende kunnskapsgrunnlag eller motstridende informasjon. Ved å benytte formaliserte konsensusprosesser får en dokumentert hva man er enig om og hva man er uenig om. Dokumentasjon om vurdering av datatilgang og validitet blir dokumentert og er dermed synlig for ettertiden.

Metodikken forutsetter et profesjonelt forarbeid av en tilrettelegger som administrerer og dokumenterer prosessene. Ulempen er at det er relativt tid- og resursskrevende å gjennomføre formaliserte konsensusprosesser. Fordelen er at det ligger et systematisk og grundig arbeid til grunn for indikatorer som skal gi uttrykk for pasienter sikkerhet. Valget er dokumentert og synliggjør hvorfor noen aspekt måles og andre aspekt ikke er valgt. På grunn av de relativt store omkostningene er metodikken best egnet til bruk av indikatorer på overordnet nivå, nasjonalt eller internasjonalt.

En kan tenke seg forenklede versjoner benyttet på lokalt nivå. Det finnes etter hvert tilgjengelig dokumentasjon av indikatorer innen de fleste medisinske og helsefaglige fagområder. Ved å bruke kun eksisterende dokumentasjon som kunnskapsgrunnlag kan man gå rett på evalueringen av enkeltindikatorer. Ved utvikling av indikatorsett bør kunnskapsgrunnlaget kvalitetssikres best mulig.

Metodikken er her beskrevet for gjennomføring i relativt små ekspertpanel, men er i andre sammenhenger også benyttet i større skala. Bruk av store panel i konsensusprosesser er kanskje bedre egnet til å identifisere hvilke områder som bør monitoreres, enn vurdering av enkeltindikatorer/indikatorsett.

I Norge har det vært liten eller ingen tradisjon for å benytte formaliserte konsensusprosesser ved valg av indikatorer innenfor helsetjenesten. Ved å kombinere kunnskap om gjennomføring av konsensusprosesser etter Rand/UCLA-metoden, med skjema fra de to europeiske initiativene, ligger forholdene til rette for å dokumentere både utvikling og valg av kvalitetsindikatorer i Norge på en god måte.

Det er mye kunnskap om utvikling av kvalitetsindikatorer, og mindre om nytten av dem i utvikling og implementering i kvalitetsforbedringsstrategier. Ved å systematisk vurdere og kvalitetssikre indikatorer i forhold til den konteksten de skal benyttes i, tror vi at kvalitetsindikatorer er et nyttig bidrag i kvalitetsforbedringsprosesser.

Notatet er et resultat av arbeidsgruppens arbeid, og anbefalingene er uttrykk for synspunktene til arbeidsgruppens medlemmer.

Key messages

Formalized consensus processes:

- Directions for difficult decision-making?

Indicators are one of many sources to knowledge about patient safety. Which indicators we choose to use, determine the picture we are painting. How these choices are made and documented, are important sources of information when the results of the indicators shall be interpreted and comprehended.

Many nations and international organizations use one or another form for formalized consensus processes in the development and choice of quality indicators for the health services. In this report we examine the principle behind the formalized consensus processes and describes one American method and two European tools which have been used for this purpose.

In Norway, there are little or no tradition for conducting formalized consensus processes when choosing indicators within the health services. By building on the knowledge from other land and organizations within this field, we can implant this method by relative simple means. Developing medical quality registries and national indicators can profit from implementing formal consensus processes. There will be an extra cost by introducing this technique, but in the long term, the knowledge base will be sounder and less time used on developing and validating the indicators.

Documentation of consensus processes provide extensive information about each single indicator and how it is considered within the medical environment. Strength and weaknesses for the indicators are documented, and are important background information for interpreting the results when the indicators are used. High cost for data collection may lead to exclusion of indicators, this is important knowledge to get the whole picture of information about sets of indicators. Transparency about indicator selection and the consequences of them are important factors in improving the understanding and use of quality indicators for the health services.

Executive summary

Formalized consensus processes:

- Directions for difficult decision-making?

BACKGROUND

Knowledge about patient safety arises from many sources. Patient safety indicators may provide systematic knowledge about certain aspects of Patient Safety. Before selecting indicators several dimensions should be considered: How to choose indicators? Who will choose the indicators to be implemented? Which criteria for documentation should give directions for the decision-making processes?

This report describes three different consensus processes and discusses what is best suited for choosing patient safety indicators for Norway.

Formalized consensus process

Formalized consensus process can in short be described as a three stage evaluation process. Characteristic of the standardized consensus process is a systematic approach, with a clear description of standards for the documentation of the knowledge base and criteria to for the evaluation, followed by documentation of both processes and outcomes.

First step in a consensus processes is identification and obtaining documentation for the indicator. Professional researchers with a competence in systematic literature review will normally be in charge of providing an overview of existing relevant evidence, methods and standards. Consensus processes will often also include studies that do not fulfil the same strict method criteria as evidence based literature reviews do.

A panel of experts evaluates validity and feasibility of the indicator based on the available documentation. Access to data is considered together with validity and reliability for each indicator. The burden of data collection must be considered alongside the utility value of the indicator. This evaluation is the second stage of the consensus process.

Composition of the expert group is an important factor. The choice of experts can be narrow within one medical speciality or area of health science, or wider, where experts are chosen from different medical specialities and areas of health science.

Expert groups rarely meet at the very first evaluation of the indicators. Documents are distributed to the experts together with a description of the further process for the evaluation of the indicator. The documentation should include a systematic literature review, list of indicators, definition of indicators and instructions. The evaluation is conducted by a survey whereupon indicators are scored based and different measures of validity and access to data specified for the indicator.

The third and last stage in the process is to gather the group of experts to follow through the final evaluation of the indicators. First is a gathering of the professionals with a presentation of the results from the survey, followed by plenary discussions and a new questionnaire evaluation of the indicators. The purpose of the consensus process is not to force a common consensus, but to document the degree of agreement/disagreement regarding the evaluation of the indicators. Indicators that reach a specified level of agreement may be chosen and indicators with a clear degree of disagreement are excluded.

One American method and two European tools

This report provides an introduction to one American method for conducting consensus processes. It is the Rand/UCLA Appropriateness Method which is recognized in work conducted for the Agency for Healthcare Research and Quality in the US, and also is at the base for the formal consensus processes in the National Quality Forum (NQF) in the US. The two European tools are Stepwise Assessment Indicator Framework Approach (SAIFA) which have been presented and used in a Patient Safety project in Europe, and the other is Appraisal of Indicators through Research and Evaluation (AIRE) developed in the Netherlands. These tools are chosen on the basis that they are specially directed towards the area of indicator selection and evaluation.

Area of interest

Within health services research and administration, formalized consensus processes are conducted when there is a weaker knowledgebase or contradictory information within a field/area. Conducting formalized consensus processes provide a documentation of agreement as well as disagreement. Documentation on the evaluation of data access and validity are valuable information for the future.

The methods imply a professional preparation of a conductor, which administrates and documents the processes. The disadvantage of conducting these processes is that they are relative time and resource demanding. The advantage is the systematic and thorough work as a base for indicators that should provide information about

patient safety. The choice is documented and make visible why some aspects are chosen and others not. Because of the relative resource demanding method, it is best suited for selection of indicators on higher levels, national or international.

One solution is to implement simplified versions on a local level. There are available documentation of indicators within most medical specialities or health science areas. By relying solely on already available documentation one can skip the initial phases and move straight to the evaluation of individual indicators. When developing sets of indicators the knowledge base should be properly quality assured.

The method is described for small expert groups, but is in some settings used for large-scale assessment. Use of large-scale assessment may be considered better suited for identifying areas for monitoring, rather than used for evaluating single indicators or sets of indicators.

In Norway, there are little or no tradition for conducting formalized consensus processes when choosing indicators within the health services. By combining knowledge of the methods for consensus processes based on the Rand/UCLA Appropriate method with the tools from the two European initiatives, we have a basis for documenting both development and choice of patient safety indicators for Norway in a proper way.

There is a lot of information about development of indicators, and less of the effect of then in development and implementation strategies for quality improvement. We believe that by a systematic evaluation and quality assurance of the indicators in relation to the context they operate within, they are valuable contributions to quality improvement strategies.

This work is a product of an independent working group and the views presented are the views of this group.

Norwegian Knowledge Centre for the Health Services summarizes and disseminates evidence concerning the effect of treatments, methods, and interventions in health services, in addition to monitoring health service quality. Our goal is to support good decision making in order to provide patients in Norway with the best possible care. The Centre is organized under The Directorate of Health, but is scientifically and professionally independent. The Centre has no authority to develop health policy or responsibility to implement policies.

Norwegian Knowledge Centre for the Health Services
PB 7004 St. Olavs plass
N-0130 Oslo, Norway
Telephone: +47 23 25 50 00
E-mail: post@kunnskapssenteret.no
Full report (pdf): www.kunnskapssenteret.no

Innhold

FORORD	11
PROBLEMSTILLING	12
1. INNLEDNING	13
Konsensusprosesser: metoder og verktøy	14
Bruk av konsensusprosesser	16
Notatets oppbygging og avgrensing	17
2. RAND/UCLA-METODIKK	18
Forarbeid	19
Rangeringer	21
Evaluering og etterarbeid	24
Vurdering av Rand/Uclametodikken	26
3. STEPWISE ASSESSMENT INDICATOR FRAMEWORK	
APPROACH	27
Forarbeid	27
Rangeringer	29
Evaluering og etterarbeid	30
Vurdering av SAIFA	31
4. APPRAISAL OF INDICATORS THROUGH RESEARCH AND	
EVALUATION (AIRE)	32
Forarbeid	32
Rangeringer	33
Evaluering og etterarbeid	34
Vurdering av AIRE	34
SAMMENFATNING	35
Beskrivelse og vurdering	35
Momenter til diskusjon	37
Arbeidsgruppens Anbefalinger	38
LITTERATURLISTE	39

Forord

Norge deltar i OECD sitt pilotprosjekt om overordna kvalitetsindikatorer for helse-tjenesten, og Nasjonalt kunnskapssenter for helsetjenesten hadde ansvar for beregning av indikatorene for pasientsikkerhet i 2008.

Kvalitetsindikatorer og ulike styringsindikatorer brukes i økende grad for å beskrive helsetjenesten. Dokumentasjon av valg og utvikling av kvalitetsindikatorer er i den sammenhengen viktig. En rekke organisasjoner og land som OECD, WHO, EU, USA, Canada, England, Nederland, Danmark og Sverige benytter alle ulike varianter av formaliserte konsensusprosesser. Deltakelse i OECD sitt kvalitetsindikatorprosjekt aktualiserte spørsmålet om indikatorene ville være egnet til nasjonal bruk også.

Initiativ til en arbeidsgruppe som skulle se nærmere på denne problemstillingen ble tatt av Enhet for pasientsikkerhet på rundebordskonferansen i november 2007. Formålet med arbeidsgruppen sitt arbeid er å se nærmere på hvordan ulike konsensusprosesser er dokumentert andre steder, vurdere ulike tilnærminger og komme med forslag til hvordan en slik prosess kan gjennomføres i Norge.

Arbeidsgruppen har bestått av:

- Prosjektleder og koordinator: forsker, Ingrid Sperre Saunes, Kunnskapssenteret
- Ellen Deilkås, Den norske legeforening, HØKH
- Marit Kristoffersen, Sykehuset Asker og Bærum
- Merete Lyngstad, Norsk Sykepleierforbund
- Bjørnar Nyen, seniorrådgiver, Nasjonalt kunnskapssenter for helsetjenesten

Problemstilling

Pasientsikkerhetsindikatorer er en kilde til kunnskap om pasientsikkerhet. Hvilke indikatorer som velges har betydning for hvilke sider ved pasientsikkerhet vi får kunnskap om. Når indikatorer brukes er det derfor flere forhold som har betydning: Hvordan velges indikatorene? Hvem velger hvilke indikatorer som skal benyttes? Hvilket krav til dokumentasjon skal det være i prosessene som fører fram til valget?

Formålet med arbeidsgruppen sitt arbeid er å se nærmere på hvordan ulike konsensusprosesser er dokumentert andre steder, beskrive de ulike tilnærminger, og komme med forslag til hvordan en slik prosess kan gjennomføres i Norge.

Ingrid Sperre Saunes

Ellen Deilkås

Marit Kristoffersen

Merete Lyngstad

Bjørnar Nyen

1. Innledning

Kunnskap om pasientsikkerhet kommer fra mange ulike kilder. Både kvalitative og kvantitative metoder må benyttes for å få et nyansert bilde(1-4). Kvalitative analyser som kjerneårsaksanalyse, dokumentanalyse, journalgjennomgang og observasjonsstudier er metoder som gir et rikholdig bilde av hvordan pasienters sikkerhet blir ivaretatt i møte med helsetjenesten. Kvantitative metoder innenfor pasientsikkerhet omfatter bruken av indikatorer og epidemiologiske studier. Det gir systematisk kunnskap om noen aspekt ved pasientsikkerhet. Kvantitative data gjør det mulig å sammenligne utvikling på enkeltområder over tid og også mellom institusjoner/enheter, og de kan benyttes for å måle effekt av intervensjoner(5).

Når indikatorer skal utvikles kan man enten velge systematiske eller usystematiske metoder i framstillingen. Usystematiske metoder baserer seg ofte på tilgjengelige data og/eller enkeltstående kritiske hendelser(6). Systematisk utvikling av indikatorer baserer seg på en gjennomgang kunnskapsgrunnlaget og kriterier som indikatorer må oppfylle for å kunne tas i bruk. Innenfor helsetjenesteforskning og administrasjon er formaliserte konsensusprosesser en metode for å bidra til beslutninger ved manglende kunnskapsgrunnlag eller motstridende informasjon.

Kvantitative metoder som kunnskapsoppsummeringer og metaanalyser er utviklet for å gi en statistisk oversikt over publiserte kliniske studier og for å få dokumentert eventuelle uoverensstemmelser i de publiserte studiene. Konsensusmetoder er en annen metode som kan benyttes for å oppnå det samme. Forskjellen er at konsensusprosesser tillater inklusjon av flere typer studier enn tradisjonelt statistiske oversikter, samt at kvalitative vurderinger av evidensgrunnlaget har en større rolle(7).

Formaliserte konsensusprosesser kan kort beskrives som en tre-trinns evalueringsprosess. I første trinn er det et forberedende arbeid med å identifiseres og framskaffes dokumentasjon for tiltak/indikator. Deretter vurderer deltakerne i et ekspertpanel hver for seg dokumentasjonen med hensyn til validitet og antatt egnethet og anvendbarhet av tiltaket/indikatoren. Dette analyseres og resultatet presenteres for ekspertpanelet som i det tredje trinnet rangerer tiltakene/indikatorerne på nytt. Metodikken forutsetter et profesjonelt forarbeid av en tilrettelegger som administrerer og dokumenterer prosessene.

KONSENSUSPROSESSER: METODER OG VERKTØY

Konsensus betyr enighet, og *konsensusstudier* undersøker hvordan enighet oppstår og oppnås. Formelle prosesser kan benyttes for å organisere både hvem som skal delta i prosessene, hva man skal fokusere på, og i hvilken grad enighet oppnås(8). Metoden har røtter tilbake til antikken og det gamle Roma. Valget av pave i den katolske kirke er kanskje den best kjente konsensusprosessen. Når en pave dør stenges kardinalene inne i en konklave til de oppnår enighet om hvem som skal bli ny pave. Det er bestemmelser for hvem som får delta i prosessen, kriterier som må være oppfylt for å kunne velges, regler for avstemming, og grad av enighet.

Hvordan prosessene organiseres kan og vil variere med formålet til den aktuelle prosessen. Her omtales kort fire forskjellige utforminger av konsensusprosesser som kan knyttes til valg av indikatorer i helsetjenesten, og to skåringsverktøy for utvikling og valg av indikatorer.

Delphi-prosesser

Fellesbetegnelse på konsensusprosesser hvor det forutsettes at deltakerne *ikke* møtes fysisk. Det er ingen direkte ansikt til ansikts interaksjon. Hovedprinsippet er at det er en gruppe/fagmiljø fungerer som tilrettelegger for konsensusprosesser hvor man søker å dokumentere grad av uenighet innenfor utvalgte områder.

Utgangspunktet er et samarbeid mellom tilretteleggere og en rådgivende ekspertgruppe. Ekspertgruppen er valgt ut av tilretteleggerne, og er satt sammen av ulike brukergrupper/interessenter på et område. Samarbeidet skal resultere i et knippe spørsmål eller påstander i form av et spørreskjema for det området/feltet som skal undersøkes.

Formålet med spørreskjemaet kan for eksempel være rangering og vurdering av validitet av tiltak/indikator. Skjemaet sendes ut til og besvares av eksterne eksperter, før det analyseres og svarene struktureres etter et gitt mønster. Dette er første runde i en Delphi-prosess.

Deretter sendes samme skjema ut på nytt til ekspertene som har svart, men nå sammen med resultatene fra første runde i delphi-prosessen. Den enkelte svarer får da vite gjennomsnittsskår, spredningsmål og egne skåringer for hvert enkelt spørsmål. De eksterne ekspertene blir bedt om å sette seg inn i resultatene fra første skåringsrunde før de svarer på skjemaet igjen. Ved å gjenta denne prosessen 2-3 runder, identifiseres de områdene hvor det oppnår konsensus.

Det finnes flere ulike varianter av Delphi-prosesser, hvor fellestrekket er at skårerens anonymitet bevares. Formålet er å unngå at gruppedynamikk får en selvstendig effekt på skåringsresultater. utfordringer for tilretteleggerne er både valg av spørsmålsstilling og utvelgelse av indikatorene som vurderes. Videre bør det være klare kriterier ved valg av ekspertgrupper. Delphi-prosesser er en betegnelse på metodikk for konsensusprosesser med tilhørende skjema utvikling.

Nominal Group Technique

Nominal Group Technique (NGT) er en form for konsensusprosess som også er klart strukturert, og hvor deltakerne møtes. Den kan benyttes i grupper av ulike størrelser, og har til formål å sikre at ulike meninger blir hørt og vurdert. Deltakerne møtes og skriver ned alle sine ideer om et gitt tema, deretter skal hver deltaker gi en kort muntlig beskrivelse av sin viktigste ide. Når alle har fått presentert sin ide diskuteres de ulike emnene. Deltakerne skårer deretter hvert tema på en numerisk skala som angir hvor viktig tema er. Totalskår angir hvilket tema som får størst oppslutning.

Metoden har flere mer eller mindre, kjente varianter som er mindre strukturerte. Idemyldring, eller brainstorming, kan sees som en avart av NGT. Deltakere inviteres til å komme med innspill som noteres ned, ideer sorteres og kategoriseres, før de stemmes over. Ekspertgruppemøter vil ofte sortere under denne kategorien(9,10).

Rand/Ucla Metodikk

Rand/Ucla-metodikk (Rand/Ucla appropriateness method - RAM) er en systematisk utvikling av retningslinjer for behandling. Metoden har i økende grad også blitt benyttet for utvikling av og valg av kvalitetsindikatorer. Hovedprinsippet er å koble evidensbasert kunnskapsgrunnlag med ekspertuttalelser/meninger for å gi retningslinjer for best mulig praksis(8,11).

Metodikken forutsetter et forarbeid av en tilrettelegger som administrerer og dokumenterer prosessene. Utgangspunktet er en systematisk litteraturoversikt som gjennomgås av ulike eksperter på området. På grunnlag av litteraturgjennomgangen samt konsultasjon med ekspertene, blir et foreløpig sett av indikatorer anbefalt av tilrettelegger (som regel forfatter/prosjektleder).

En ekstern ekspertgruppe rekrutteres, og de rangerer kvalitetsindikatorerne etter gitte kriterier. Indikatorerne rangeres i to omganger: første gang sendes spørreskjema postalt, deretter analyseres svarene før ekspertgruppen samles. Når ekspertene er samlet diskuteres resultatene før en skårer indikatorerne på nytt.

Rand/Ucla-metodikk er for konsensusprosesser som baseres på systematisk kunnskapsoppsummering og tilhørende skjema utvikling.

NIH Consensus Conference

Konsensuskonferansene som U.S. National Institutes of Health arrangerer, er en variant av konsensuskonferanser hvor tema utredes på forhånd, representanter for ulike grupper møtes og blir enig om en felles uttalelse. Før en konsensuskonferanse gjennomføres det en systematisk litteraturgjennomgang og oppsummering av kunnskap. Temaene presenteres av forskere og ulike talsmenn, og de diskuteres i plenumsdiskusjoner. Konferansene varer 2 eller flere dager og avsluttes ikke før deltakerne har blitt enige om en skriftlig uttalelse. Deltakerne er leger, forskere og representanter for brukerne av tjenestene. Formålet med konferansen er å få svar på spesifikke, forhåndsdefinerte spørsmål.(12)

Stepwise Assessment Indicator Framework Approach (SAIFA)

Stepwise Assessment Indicator Framework Approach (SAIFA) er et verktøy utviklet av European Society for Quality in Healthcare (ESQH). Det er en trinnvis prosess som er laget for å karakterisere og evaluere indikatorer for et pasientsikkerhetsprosjekt i EU(13).

SAIFA har klare likhetstrekk med Rand/Ucla-metodikken. En ekspertgruppe har utviklet et eget skjema for å beskrive hver enkelt indikator, beskrivelsen viser også til dokumentasjonsgrunnlaget for den enkelte indikator. Det er også utviklet et eget skjema for vurdering av indikatorene. Det brukes for å rangere indikatorene i forhold til de tre dimensjonene: relevans, validitet/reliabilitet og anvendelighet.

Metode for gjennomføring av konsensusprosesser er ikke systematisk beskrevet. SAIFA framstår derfor mer som et verktøy for indikatorbeskrivelse og rangering enn metode for vurdering.

Appraisal of Indicators through research and Evaluation (AIRE)

Appraisal of Indicators through Research and Evaluation (AIRE) er et instrument utviklet i Nederland for **vurdering** av kvalitetsindikatorer. Det skiller seg dermed noe ut fra de andre instrumentene ved at det ikke primært er konstruert for valg av indikatorer, men alle har til felles at de vurderer det vitenskaplige grunnlaget for indikatorer. AIRE vurderer i tillegg formål, relevans og organisasjonskontekst for utvikling av indikatorene, samt interessenters involvering i utvikling av indikatorer. Fokus settes dermed også på konteksten til indikatorene. Indikatorbeskrivelsen kan også benyttes som en sjekklister ved utvikling av kvalitetsindikatorer(14). Verktøyet kan dermed nyttes både ved vurdering og utvikling av kvalitetsindikatorer.

Vi har ikke informasjon om hvordan skåringsprosessen styres. Skjemaet er under utvikling, og skåringsskjemaet er tilgjengelig. Informasjonsmengden i skjemaet kan tyde på at det er tenkt gjennomført som Delphi-prosesser. AIRE framstår nå som et vurderingsskjema/verktøy tenkt benyttet i konsensusprosesser.

BRUK AV KONSENSUSPROSESSER

Ulike varianter av denne metodiske tilnærmingen er benyttet i flere land og internasjonale organisasjoner for å velge kvalitetsindikatorer(6,8,9,11,15). Et sentralt argument for å benytte konsensusprosesser som arbeidsmetode ved utvikling av kvalitetsindikatorer er at evidensgrunnlaget for valg av kvalitetsindikatorer er svakt. Ved å systematisk oppsummere kunnskapsgrunnlaget og benytte det for evaluering og valg av indikatorer kan en dokumentere hvilke valg som er tatt, grunnlaget for valgene og eventuell uenighet om resultatet.

En kvalitetsindikator kan ha god evidens som indikator på en egenskap ved kvalitet. Evidensgrunnlaget for å velge den indikatoren framfor en annen er likevel avhengig av hvilke andre indikatorer som den presenteres i lag med. Evidensgrunnlaget til

kvalitetsindikatoren må derfor sees i sammenheng med evidensgrunnlaget til de andre kvalitetsindikatorerne som velges. Denne koblingen mellom evidensgrunnlag og en ekspertvurdering er kjernen i de konsensusprosessene som omtales her. Konsensusprosesser benyttes derfor for å komme til enighet om hvilke indikatorer som det oppnås størst oppslutning om.

Formålet er ikke å framtvinge konsensus, men å identifisere indikatorer som kan velges og hvilke som kan utelukkes. Ved å benytte formaliserte konsensusprosesser får en dokumentert grad enighet. Vurdering av datatilgang og validitet blir også dokumentert og synlig for ettertiden.

Metodikken er her beskrevet i forhold til relativt små ekspertpanel, den kan også benyttes i større skala. Den er da kanskje bedre egnet til å identifisere hvilke områder som skal monitoreres, framfor vurdering av enkeltindikatorer/indikatorsett.

NOTATETS OPPBYGGING OG AVGRENSING

I dette notatet har vi valgt å beskrive en metode for å gjennomføre konsensusprosesser i detalj, deretter presenteres to andre metoder som også benyttes for å velge kvalitetsindikatorer i helsetjenesten.

Vi beskriver først relativt detaljert, oppbygging og gjennomføring av Rand/UCLA sin metodikk for konsensusprosesser. Agency for Healthcare Research and Quality har benyttet denne metoden, og den ligger også til grunn for gjennomføring av konsensusprosesser i National Quality Forum (NQF) i USA. Sistnevnte organisasjonen er en ideell organisasjon basert på et fellesskap av private og offentlige helsetjenesteaktører. Målsettingen til NQF er å bedre kvaliteten på helsetjenesten gjennom oppslutning om konsensusbaserte standarder for målinger og offentlige rapporteringer av ytelsesindikatorer. Vi har valgt å presentere Rand/UCLA sin metodikk ettersom den er best beskrevet og omfatter både Delphi-prosesser og ekspertgruppemøter.

Vi har også valgt å presentere to verktøy, henholdsvis SAIFA (Stepwise Assessment Indicator Framework Approach som er benyttet i EU og AIRE (Appraisal of Indicators through Research and Evaluation) som er utviklet i Nederland. Begge skjemaene er tilgjengelig på internett, og det er tilgang på skjema som er begrunnelsen for at de er tatt med her.

Avslutningsvis oppsummerer vi og foreslår framgangsmåte for bruk av konsensusprosesser ved valg av indikatorer for pasientsikkerhet i Norge.

2. Rand/UCLA-metodikk

RAND/UCLA-metodikk (RAND/UCLA appropriateness method - RAM) er en metode utviklet for å sikre best mulig beslutningsgrunnlag for klinikere/praktikere ved systematisk utvikling av retningslinjer for behandling. Den ble utviklet på midten av 80-tallet, primært som et middel for å måle under- og overforbruk av medisinske og kirurgiske prosedyrer (16). Hovedprinsippet er å koble evidensbasert kunnskapsgrunnlag med ekspertuttalelser/meninger for å gi retningslinjer for best mulig praksis.

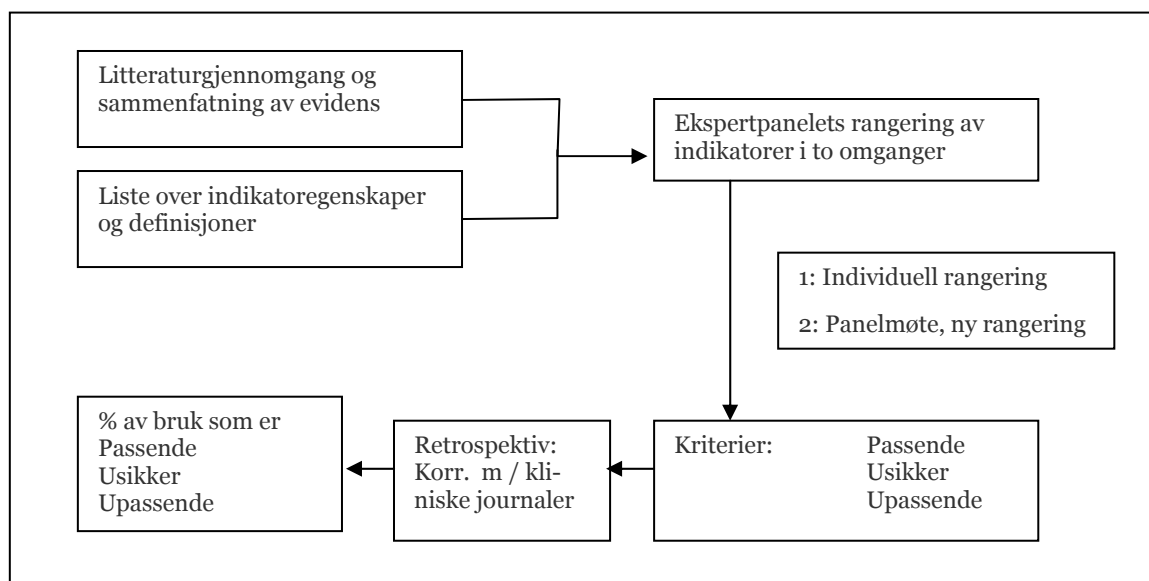
Rasjonale bak RAM var at randomiserte kontrollerte studier (RCT) ofte ikke var tilgjengelig eller populasjonen var for avgrenset til at man fikk vitenskaplig grunnlag for medisinske beslutninger og valg. Ved å utvikle denne metodikken spesifikt for helsesektoren fikk de et verktøy som kombinerer den best tilgjengelige kunnskapen med medisinsk fagekspertise. I begynnelsen var det en retrospektiv metode, men har siden utviklet seg til også å bli prospektiv. RAM er skjematisk framstilt i Figur 2.1

Metodikken forutsetter et forarbeid av en tilrettelegger som administrerer og dokumenterer prosessene. Utgangspunktet er en systematisk litteraturoversikt som gjennomgås av ulike eksperter på området. På grunnlag av litteraturgjennomgangen samt konsultasjon med ekspertene, blir et foreløpig sett av indikatorer anbefalt av forfatter/prosjektleder.

Spesialister på området rekrutteres og inviteres til å delta i ekspertpanel i en to-trinns prosess for rangering av indikatorer. Foreslåtte indikatorer og litteraturgjennomgang sendes til deltakerne i ekspertpanelet. De rangerer indikatorene med hensyn til deres validitet som kvalitetsmål og mulighetene for å samle inn de data som er spesifisert i indikatoren. Tilgjengeligheten på dataene vurderes ut fra hvor sannsynlig det er å finne opplysningene i en pasientjournal/eksisterende databaser. Paneldeltakerne gir hver indikator derfor to skårer på en kontinuerlig skala fra 1-9. En skår for validitet og en skår for tilgjengelighet.

Resultatene fra den første skåringen av indikatorene presenteres deretter for ekspertpanelet på et todagers fellesmøte. Hver enkelt deltaker får vite sin opprinnelige skåre, mål for sentraltendens og spredning, samt frekvensfordelingen av skåringene. Alle indikatorene diskuteres, modifiseres om nødvendig, og skåres på nytt. Siste skåringsrunde blir benyttet for å velge ut de indikatorene som skårer høyt på validitet og tilgjengelighet hos paneldeltakerne.

Figur 2.1 Skjematisk oversikt over RAM



Beskrivelse av prosessene er basert på framstillingen i manualen til RAM.

FORARBEID

Vurdering og sammenfatning av litteratur

Metodikken forutsetter et forarbeid av en tilrettelegger som administrerer og dokumenterer prosessene. Arbeidet starter med en systematisk litteraturoversikt som gjennomgås av ulike eksperter på området. Litteraturoversikten skal være tilgjengelig for alle som deltar i prosessen slik at de har felles kunnskapsgrunnlag. Den skal også fungere som en ressurs for eventuelle avklaringer, da den har referanser til aktuelle studier.

Litteraturoversikten skal inneholde søkemetodene og strategiene som er benyttet. Det anbefales også at man sjekker referanselistene til inkludert litteratur for å sikre at man har fått med seg relevant informasjon. Sammendrag av ikke publiserte studier ekskluderes (jfr posterpresentasjoner). Eksperter kan også konsulteres for å sikre at man får med tilgjengelige studier. Svært sensitive søkestrategier er mindre spesifikke, og det er ofte behov for å begrense litteraturtilfanget. Avgrensingene kan knyttes til språk, inklusjonsperiode etc. Klassifisering av litteratur er ofte knyttet til metode, hvor store randomiserte kontrollerte studier er en kategori, mindre randomiserte studier en annen og så videre.

Beskrivelse av indikator

Oppsummering av litteratur sammenfattes i en beskrivelse av den enkelte indikator. En sammenfatning av egenskaper ved pasientsikkerhetsindikatorer som skal dokumenteres etter RAM er blant annet gitt i en dokumentasjonsrapport til OECD:

- Indikatorens relevans/betydning for pasientsikkerhet
- Vitenskaplig dokumentasjon
- Potensial med hensyn til gjennomførbarhet og egnethet

Indikatoren for pasientsikkerhet skal dekke forhold som kan forebygges, eller er knyttet til feil ved behandling eller medisinerings. Indikatorens relevans, eller innholdsvaliditet må beskrives i forhold til kunnskapsgrunlaget for pasientsikkerhet.

Den vitenskaplige dokumentasjonen har hovedvekt på begrepsvaliditet, også omtalt som konstruksjonsvaliditet. En indikator for pasientsikkerhet vurderes opp mot andre mål for pasientsikkerhet som for eksempel dødelighet, reinnleggelse, forlenget oppholdstid eller varige funksjonsnedsettelse hos pasienter. Det er relativt vanlig å sammenholde prosessmål med ulike resultatmål, men resultatmål vurderes også i forhold til resultatmål(17). Vitenskaplig dokumentasjon av kriterievaliditet er også viktig. En sammenligner da resultatene av indikatormålinger med en "gullstandard" om det eksisterer, eller med mer komplekse målinger av det samme fenomenet. Resultater fra indikatormålinger i pasientadministrative system kan for eksempel valideres i forhold til journalgjennomganger.

Vurderinger i litteraturen av indikatorens potensial til å føre til forbedring, skal dokumenteres. Under dette punktet sammenfattes kunnskap om tilgang på data og omkostninger ved datainnsamling på ulike nivå. Datakvalitet er også et viktig moment i denne sammenhengen.

Det viktigste i denne fasen er å komme fram til så klare definisjoner som mulig for å oppnå best mulig resultat. På grunnlag av litteraturgjennomgangen, samt konsultasjon med ekspertene, blir et foreløpig sett av indikatorer anbefalt av forfatter/prosjektleder.

Ekspertpanelet

Ekspertpanelet skal i en totrinns prosess rangere indikatorene med hensyn til validitet og tilgjengelighet. Sammensetning av ekspertpanelet er svært viktig. Enkelte panel har valgt kun å benytte eksperter innenfor et medisinsk område, men de fleste er bredere sammensatt(8;10;16). De består gjerne både av utøvere innenfor et klinisk område, klinikere i relaterte fagområder og eventuelt allmennleger. Studier av panelsammensetning har vist at spesialister har en tendens til å rangere prosedyrer eller tiltak innenfor egen spesialitet høyere enn prosedyrer som ikke tilligger egen spesialitet. De fleste RAND/UCLA panel er bredt sammensatt for bedre å representere variasjonen av spesialiteter som faktisk er involvert i beslutninger om hvilke tiltak som skal iverksettes(8).

Ettersom det er nasjonale variasjoner i hvem som er involvert i valg av indikatorer, vil det være nasjonale variasjoner i hvilke anbefalinger man kommer med for sammensettingen av ekspertpanel. I USA anbefales det at spesialistforeninger inviteres til å foreslå paneldeltakere, og at universitetsansatte og enkeltpersoner inviteres med. Ved å gå bredt ut i rekrutteringsprosessen til ekspertpanelet får man en større utvalgsbase å velge fra. Er utvalget av kandidater til ekspertgruppen stort, kan en kvasirandomisere utvelgelsen. Hovedkriterier for å bli valgt er å ha en anerkjent posisjon innenfor sin spesialitet, fravær av motstridende interesser i fagfeltet, geografisk spredning og mangfold i praksiserfaringer. Medlemmer bør ikke velges kun fordi de er tilgjengelige, eller for å holde reisekostnader på et lavt nivå.

De tidligste Rand/Ucla panel var sammensatt av 9 medlemmer, men andre studier har benyttet fra 7-15 medlemmer. Den opprinnelige valideringen av metodikken for konsensusprosesser er beregnet ut fra 9 individ, og det har blitt en svært utbredt gruppestørrelse. Det er viktig at man tar hensyn til representativitet når man skal sette sammen et panel, men det frarådes å ha mer enn 15 medlemmer i panelet.

Motivering av paneldeltakere er stort sett knyttet til å betale utgifter forbundet med deltakelse. Enkelte gir en symbolsk betaling, og europeiske erfaringer tyder på at ekspertene deltar uten økonomisk belønning.

RANGERINGER

Når paneldokumentene er klare, sendes de til deltakerne med en forklaring av videre saksgang. Paneldokumentene skal inneholde litteraturgjennomgang, liste over indikatorer, definisjoner og instruksjoner. Det er anbefalt å følge opp utsendelsen med en telefonsamtale for å forsikre seg om at dokumentene har kommet fram, og at instruksene er klare. Når paneldeltakerne sender inn svar, bekreftes mottakelsen av innsendingen. Gjennom rangeringsrundene skal panelet komme fram til en felles vurdering av indikatorer som enten passende, usikker eller upassende. Fastsetting av kriterier for å avgjøre om en indikator er passende, involverer minst to runder av rangeringer, en rangering før deltakerne møtes og en rangering under møtet.

Ekspertpanelet blir bedt om å vurdere egnetheten/nyttens av indikatorene på en skala fra 1 til 9, hvor 1 angir svært lite egnet, og 9 er svært egnet.

Første rangeringsrunde

Paneldeltakerne får tilsendt litteraturoppsummering og annen nødvendig informasjon for å rangere prosedyrene/indikatorene. Svarfrist settes til 4-6 uker. En beregner at paneldeltakerne trenger 5-10 timer for gjennomlesning, avhengig av omfang på litteratur. Paneldeltakerne vet ofte ikke hvem de andre ekspertene i panelet er om det ikke er avholdt prekonferanser.

Enkelte ganger blir paneldeltakerens bedt om å vurdere av kunnskapsgrunnlaget for indikatorene. En ber dermed både om ekspertens profesjonelle vurdering av indikatoren og om hvordan eksperten vurderer det kunnskapsgrunnlaget som er blitt presentert for indikatoren.

I løpet av dette tidsrommet kan deltakerne i panelet kontaktes slik at de får anledning til å kommentere innholdet. Eventuelle kommentarer samles og diskuteres på fellesmøte.

Andre rangeringsrunde

Resultatene fra første rangeringsrunde oppsummeres og presenteres for deltakerne på konsensuskonferansen. Hver deltaker skal ha tilbakemelding på frekvensfordelinger for hver indikator, inkludert egne rangeringer. Dette dokumentet danner grunnlag for diskusjonene på konsensusmøtet (se fig 2.2).

Figur 2.2: Presentasjon av rangering for deltaker.

Indikator xxx	6	2	2	1	2	1				1
	1 ^{*a}	2	3	4	5	6	7	8	9	(2.0) ^b (1.8) ^c

Øverste linje: frekvensfordeling

Nederste linje: svarskala

a: paneldeltakerens egen skåring markert med asterix

b: median

c: standardavviket til median

I løpet av konsensusmøtet diskuteres resultatene fra første rangeringsrunde. Alle paneldeltakerne har da tilgjengelig dokumentasjon som viser hvordan de rangerte indikasjonen på forhånd, samtidig som de ser hvordan panelet som helhet fordeler seg på rangeringen. Dokumentasjonen gir innspill til diskusjonen ved at den oppsummerer alle deltakernes syn på hver enkelt indikator uten at informasjonen knyttes til enkeltindivider. Det er en effektiv presentasjon av synspunkt hvor man raskt får oversikt over hvor stor eller liten spredning det er blant deltakerne for det enkelte diskusjonspunkt. Alle indikatorene diskuteres, modifiseres om nødvendig, og skåres på nytt. Siste skåringsrunde blir benyttet for å velge ut de indikatorene som skårer høyt på validitet og tilgjengelighet hos paneldeltakerne.

Til å lede konsensusmøtet er det valgt en moderator som tilhører arbeidsgruppen som har ansvar for å forberede og gjennomføre konsensusprosessen. Moderator/møteleder benytter tilsvarende oppsummering som paneldeltakerne, men i noe modifisert utgave (Figur 2.3).

Figur 2.3: Presentasjon av rangeringsfordeling for moderators bruk.

Indikator x	6	2	2	1	2	1	1	Median	SD	Enighet			
	1	2	3	4	5	6	7	8	9	(2.0) ^a	(1.8) ^b	A ^c	- ^d

Øverste linje: frekvensfordeling

Nederste linje: svarskala

a: median

b: standardavviket til median

c: Nivå for enighet

d: Samlet vurdering av indikatorens nytte angies ved + eller -

De to siste kolonnene som moderator har i sin presentasjon angir nivå av enighet og vurdering av hvor formålstjenlig indikatoren er. A angir at det er stor grad av enighet (**A**greement) blant paneldeltakerne i rangering av indikatoren. Minustegnet angir at de ikke vurderer indikatoren som formålstjenlig. Dette er en rask oppsummering som benyttes av moderator for å lede de videre diskusjoner. I tillegg kan moderator også benytte oversikter som viser hvordan den enkelte paneldeltaker har rangert indikatorene. Figur 2.4 er et eksempel på en slik oversikt. Merk at nederste linje her ikke er svarskala som i figur 2.2 og 2.3, men her angir paneldeltakernes nummer.

Figur 2.4: Moderators oversikt over paneldeltakeres rangering av indikatoren.

Indikator x	9	8	2	9	9	7	6	8	5	9	8	7	Median	
	1	2	3	4	5	6	7	8	9	10	11	12	8	A ^a + ^b

Nederste linje: paneldeltakerenes identitet/nummer

Øverste linje: paneldeltakerens skåring av indikator

a: Nivå for enighet

b: Samlet vurdering av indikatorens nytte angies ved + eller -

Moderators rolle

Moderator har ansvar for gjennomføring av konsensusprosessen, og bør fortrinnsvis være en ekspert på området, men ikke med egen faglig bakgrunn for nært opp mot det som skal rangeres(8,11). Dette for å forebygge at personlige preferanser ikke får innflytelse på arbeidet. Moderator bør ha god kjennskap til litteraturgjennomgangen, og fortrinnsvis ha deltatt aktivt i dette arbeidet. Moderator bør være kontaktpunktet for paneldeltakerne mellom første og andre rangeringsrunde.

Etter at oppsummeringen av første runde er klar, bør moderator forberede seg ved å markere hvilke indikatorer som er vurdert som passende, usikre og ikke passende. Det bør også markeres hvor det er inkonsistens i vurderingene slik at dette kan tas opp til nærmere diskusjon. Ved innledning til møte bør moderator oppsummere hva som er formålet med rangeringen, minne om målsettingen med vurderingen. Deretter bør man gi en rask oversikt over metodikken som benyttes, slik at alle har klart i minne hva som forventes av innsats. Det må eksplisitt minnes om at formålet er ikke

å framtvinge enighet. Videre at kostnadsvurderinger ikke må influere på deres vurderinger. Det er relevant i andre sammenhenger, men ikke når man skal vurdere om en metode er passende.

Gjennomgang av indikatorlisten for diskusjon kan deretter starte. Det anbefales å starte med relativt enkle indikatorer først for å få fortgang i prosessene. Åpner møtet med de mer vanskelige vurderingene vil det kunne forsinke prosessene. Moderator skal verken dominere diskusjonen eller la paneldeltakerne bli fordypet i mindre fruktbare diskusjoner om detaljer som ikke er vesentlig.

EVALUERING OG ETTERARBEID

Etter hver skåringsrunde sammenfattes resultatene for å få oversikt over hvordan indikatorene ble vurdert og hvordan oppslutningen om vurderingen var. Definisjonene for grad av oppslutning om en indikator, er avhengig av antall deltakere i panelgruppene.

Vurdering

Rangeringen av indikator oppsummeres i følgende kategorier: passende, usikker og upassende.

Kategoriseringen baseres på skåringsresultatene for median og spredningsmål for den. Medianen angir i hovedsak hvilken kategori en indikasjon får, og spredningsmål viser grad av enighet/uenighet som ligger til grunn for vurderingen. Man blir enig om hvilke grenser man skal ligge til grunn for kategorisering før man setter i gang skåringsprosessen.

Følgende definisjoner av kategoriene er foreslått:

Passende: median rangering på 7-9, uten avvikende skåringer

Usikker: median rangering på 4-6 / median med avvikende skåringer

Upassende: median skåring på 1-3, uten avvikende skåringer

Denne definisjonen er tilpasset enhver panelstørrelse med et oddetalls antall deltakere. Er det et partalls antall deltakere i et panel er det mulig med desimalskåringer for median. Da må det avgjøres på forhånd hvordan man skal kategorisere medianskåringer som faller mellom de ulike kategoriene, det vil si medianskårer på 3,5 og 6,5.

Oppslutning

For å avgjøre oppslutning om et standpunkt må en i utgangspunktet bli enig om kriteriene for oppslutning. Ulike tilnærminger er benyttet. Under utvikling av denne metodikken ble det benyttet ulike definisjoner for enighet, etter hvert har en lagt vekk på hva som konstituerer uenighet. Om resultatene ikke kan kategoriseres som

uenighet, er det konsensus eller oppslutning om vurderingen, enten vurderingen er at indikatoren passer, en er usikker eller at den vurderes som upassende.

Den opprinnelige definisjonen for uenighet krevde at minst en hadde skåret 1 og en annen 9. Det vil si at minst to av ni personer i ekspertgruppen må vurdere indikatoren i hvert sitt ytterpunkt av skalaen. En noe romsligere fortolkning av uenighet er at minst to av ni personer må skåre i henholdsvis det nederste (1-3) og det øverste (7-9) intervallet.

Vi gjengir her i tabell 2.1 den klassiske definisjonen for enighet/uenighet som benyttes ved RAM(8).

Tabell 2.1: Den klassiske definisjonen som Rand forskerne benytter:

Enighet	Ikke mer enn to paneldeltakere skårer indikasjonen utenfor det trepunkts intervall (1-3, 4-6, 7-9) som medianen ligger i.
Uenighet	Minst tre paneldeltakere rangerer indikasjonen i laveste intervall (1-3) og minst tre paneldeltakere rangerer indikasjonen i øverste intervall (7-9).

Beregning av antall medlemmer som må svare i de ulike kategoriene er basert på ekspertpanel med ni medlemmer. Det er gjort en flernasjonalt, europeisk tilpasning, hvor en har beregnet antall medlemmer som må skåre i ulike kategorier for å avgjøre enighet/uenighet for ulike panelstørrelser.

Tabell 2.3: Definisjoner for enighet/uenighet for ulike panelstørrelser

Panel Størrelse	Uenighet Antall paneldeltakere som skårer i hver ekstremkategori (1-3 og 7-9)	Enighet Antall paneldeltakere som skårer utenfor medianintervallet (1-3, 4-6, 7-9)
8-9-10	≥ 3	≤ 2
11-12-13	≥ 4	≤ 3
14-15-16	≥ 5	≤ 4

Ulempen med denne tilnærmingen er at den fungerer for panelgrupper hvor antall deltakere går opp i 3, men for andre gruppestørrelser vil det være intern bias. Det vil si at det blir mindre sannsynlig at det blir uenighet i den minste gruppen, og tilsvarende større sannsynlighet for uenighet i den største gruppen.

Dette har ført til utvikling av et kontinuerlig mål for spredning: MAD-M som står for *Mean Absolute Deviation from the Median*. Målet har imidlertid ikke funnet stor oppslutning selv om det er det eneste kontinuerlige målet som er utviklet. Et nytt

mål som er utviklet er IPRAS som står for Interpercentil Range Adjusted for Symmetry. Hovedprinsippet er å undersøke om interpercentil rangen for skåringsresultatene er større enn interpercentilrangen når en har justert for assymetri. Spredningsmålet tar altså utgangspunkt i relativ spredning (målt i prosent) blant skårerne. Det går ikke klart fram av rapporten hvordan målet beregnes.

Det er ikke alle indikatorvurderinger en oppnår enighet om. Da har man på forhånd bestemt grensene for når man er enig om å være uenig.

VURDERING AV RAND/UCLAMETODIKKEN

En av de største utfordringene ved RAND/UCLA-metodikken er at det er vanskelig å estimere hvor mye tid som går med til sammenfatning av litteratur og gjennomføring av prosesser. Det er avhengig av kompleksiteten til et område og omfanget av dokumentasjon på området. Det er også et klart behov for kompetanse til å gjennomføre bibliografiske søk, kompetanse på å vurdere og rangere vitenskapelige artikler, og til å sammenfatte informasjon fra ulike kilder. For områder som tidligere ikke er sammenfattet i litteraturoversikter viser erfaring at man må regne med et minimum av seks måneders innsats. I tillegg trengs det erfaringer i rollen som tilrettelegger eller moderator for konsensusprosesser.

Fordelen er en grundig og godt dokumentert gjennomgang av eksisterende forskning som er knyttet opp mot bruk av en indikator. Dokumentasjonsgrunnlaget vil også være viktig ressurs inn i det videre indikatorarbeidet og beskrivelse av resultater av målinger.

3. Stepwise Assessment Indicator Framework Approach

Stepwise Assessment Indicator Framework Approach (SAIFA) er et system for vurdering av indikatorer. Systemet er utviklet som en del av pasientsikkerhetsprosjektet Safety Improvement for Patients In Europe (SimPatIE). SimPatIE var et EU-finansiert prosjekt som søkte å sammenfatte informasjon om pasientsikkerhet i Europa og var organisert i flere delprosjekter. Instrumentet ble utviklet ved European Society for Quality in Healthcare (ESQH) i Danmark under ledelse av Jan Mainz. Beskrivelsen av instrumentet er basert på en rapport fra arbeidsgruppe 4 i SimPatIE-prosjektet(13).

Hovedvekten i ESQH-rapporten er lagt på beskrivelse og funksjon til skjemaene for indikatorbeskrivelse og indikatorutvelgelse. Formaliteter for gjennomføring av konsensusprosesser er ikke beskrevet, kun prosessen som ble brukt i det konkrete prosjektet.

FORARBEID

For at en indikator skal velges må den tilfredsstillende to kriterier: For det første må det være enighet om at indikatoren direkte eller indirekte kan benyttes for å overvåke **pasientsikkerhets**hendelser som kunne vært unngått (SimPatIE WP4, side19). Pasientsikkerhetsindikatorer defineres dermed gjennom formålet med indikatorer, som er overvåking av hendelser, direkte eller indirekte.

Det andre kriteriet er at hendelsen kan kategoriseres i forhold til:

- faktiske hendelser versus nesten hendelser
- unngåelige hendelser versus uunngåelige hendelser
- skade versus ikke skade

Definisjonen for indikatorer fungerer som kriterium som indikatorer må tilfredsstillende for å betegnes som en pasientsikkerhetsindikator. Det ble utviklet et standardisert skjema som beskriver indikatorene, samt et evalueringsskjema for hver indikator. Skjema for indikatorbeskrivelse er gjengitt i figur 3.1.

Figur 3.1 SAIFA skjema for beskrivelse av pasientsikkerhetsindikator (PSI)

PSI (Nummer) : (Tittel)	
Overordnet begrunnelse for PSI i forhold til kontekst/organisasjon	
Dimensjon	Beskrivelse
Beskrivelse av spesifikke aspekt ved pasientsikkerhet	En klar beskrivelse av spesifikke aspekter ved pasientsikkerhet, pasientpopulasjon, tjenesteleverandører, helsetjenestetype og tidsrom
Formål med PSI	Gir en konsis beskrivelse av formålet til den spesifikke indikatoren (begrunnelse)
Nivå for beskrivelse av pasientsikkerhet	Identifiserer på hvilket nivå sikkerhet blir vurdert, ie på individnivå eller på aggregert pasientnivå
Kilde(r)	Beskriver opprinnelsen til indikatoren, hvor den tidligere er brukt
Evidens for spesifikke pasientsikkerhetskriterier	Beskriver kunnskapsgrunnlaget for valg av felt/område som indikatoren dekker (hvorfor dette er et viktig PS-område)
Omfang av klinisk testing	Beskriver omfanget av testing av indikatoren (inkludert reliabilitet og/eller validitetstesting)
Evidens for klinisk bruk av retningslinjer	Beskriver omfanget av kliniske retningslinjer i bruk, og spesifikt hvilke retningslinjer som er knyttet til indikatorens målsetting.
PSI kategori	Spesifiser hvorvidt indikatoren skal benyttes på <i>Institusjonsnivå</i> <i>Innsatsområde:</i> <ul style="list-style-type: none"> Infeksjonskontroll Kirurgiske komplikasjoner Medikamenthåndtering Fødselshjelp/obstetrics Fall i sykehus Diagnosespesifikke eller andre spesifiserte PSler
Datadefinisjon	Detaljert beskrivelse av data som indikatoren er basert på
Beskrivelse av nevner	Gir den generelle spesifikasjonen av enhver klinisk komponent som danner grunnlag for inklusjon og eksklusjon av nevner (populasjon)
Beskrivelse av teller	Gir den generelle spesifikasjonen av enhver klinisk komponent som danner grunnlag for inklusjon og eksklusjon av teller
Datakilde	Identifiserer nødvendige datakilder for å implementere målingen
Identifisere institusjonell kontekst	Plassere indikatoren i en institusjonell kontekst, feks med hensyn til indikatorens relasjon til kvalitetsforbedringsarbeid og/eller akkreditering
Tjenesteområde	Klassifisering for de helsetjenesteområder som indikatoren benyttes for
Profesjonstilknytning for helse-tjenesteytelsen	Klassifisering av personell som er ansvarlig for helsetjenesten
Laveste tjenestenivå som indikatoren rettes mot	Klassifisering det laveste tjenestenivå som indikatoren er anvendt på.

Eventuelle pasient-faktorer/populasjonsvurderinger	Identifisering pasientrelaterte faktorer som har betydning for bruk av indikatoren, og typer av analytiske betraktninger som er knyttet til indikatoren. Det er egenskaper ved pasientpopulasjonen som høy risiko/sårbare grupper, andre grupperinger (alder/kohort), case-mix justeringer etc
Stratifisering i forhold til sårbare populasjoner	Beskriver populasjoner som er utsatte/ sårbare for sviktende pasientsikkerhet av den type som indikatoren beskriver
Standarder for sammenlikning	Klassifiserer typer og tidsrammer for sammenlikninger ut fra om det skal være eksterne (tverrsnitt/tidsserie) interne eller grunnleggende retningslinjer (baselinje)
Skåringer	Identifiserer metoder brukt for å skåre målingen: Kategoriske/kontinuerlige variabler, antall, frekvensfordelinger, ikke-vektede skårer/sammensatte /skalaer, rater, andeler, vektet skåre/sammenstillinger/skalaer

Indikatorene, definisjonene og kriteriene ble diskutert i telefonkonferanser til de ble enig om et felles sett av definisjoner. Endringer i indikatordefinisjoner ble kun foretatt for indikatorer som ikke hadde fast klarlagte definisjoner. SimPatIE avgrenset seg til å vurdere eksisterende indikatorer.

RANGERINGER

Det andre skjemaet som ble utviklet i SimPatIE-prosjektet er et skåringsskjema som benyttes for rangering av indikatorene (Figur 3.2). Øverste del av skjema er informasjon om de tre dimensjonene som skal legges til grunn for evaluering av den enkelte indikator. Det er korte definisjoner og oversikt over skåringsverdier. Indikatorene skal vise om den er :

- relevante og passende
- valide og reliable
- om de er gjennomførbare

For at en indikator skal regnes som relevant og passende må den være sentral innenfor ett gitt område. Området avgrenses ved å ha klar definisjon av populasjonen eller organisasjonen en ønsker å undersøke. Sentralitet avgjøres ved å vurdere hvor ofte noe opptrer (frekvens) og omfanget av skade (alvorlighetsgrad). Den første dimensjonen som skal rangeres kan sammenfattes til indikatorens betydning for pasientsikkerhet.

Den andre dimensjonen er vurdering av gyldighet og pålitelighet for måling av indikatoren. Det blir gitt en felles vurdering av indikatoren som skal sammenfatte vurdering av konstruksjonsvaliditet (måler indikatoren det den sier den skal måle), intern konsistens (fanger målingen opp det samme fenomenet ved gjentatte målinger) sensitivitet/spesifitet (indikatoren evne til å skille mellom de som ikke skal være

med i studien og de som skal være med) og reliabilitet (kildekritikk og vurdering av pålitelighet).

Den siste vurderingen er av gjennomførbarhet, om indikatoren er gjennomførbar i praksis. En skal da både vurdere tilgang på eventuelle data, samt byrden ved eventuelt merarbeid for helsepersonell ved innsamling av data. Skåringen av de tre dimensjonene skal gjøres på en skala på 1-9, men i tillegg er det et fritekst område for kommentarer.

Figur 3.2 SAIFA - skjema for vurdering av indikator.

Skåringsmatrise			
Dimensjon	Definisjon	Skåring	
Relevans og passende	Er vesentlige områder (alvorlige og hyppige) dekket med tanke på pasientsikkerhet innen det gitte området (populasjon og/eller organisasjon)?	1-3 Liten grad relevant 4-6 Middels grad relevant 7-9 Stor grad relevant	
Validitet og reliabilitet	Er instrumentet tilfredsstillende med tanke på <ul style="list-style-type: none"> - Konstruksjonsvaliditet (kunnskapsbasert) - Intern konsistens - Uttømmende/ekskluderende - Reliabilitet 	1-3 Liten grad valid 4-6 Middels grad valid 7-9 Stor grad valid	
Gjennomførbarhet	Hvordan er: <ul style="list-style-type: none"> - Tilgang på data - Klinisk rapporteringsbyrde 	1-3 Liten grad egnet 4-6 Middels grad egnet 7-9 Stor grad egnet	
Skåringsark			
Tittel (på vurdert instrument)			
Relevans og passende	Validitet og reliabilitet	Gjennomførbarhet	Tilleggs kommentarer/ generell vurdering
Skår fra 1-9	Skår fra 1-9	Skår fra 1-9	Fri tekst

Resultatene for skåringen ble formidlet både i form av mål for sentraltendens for relevans, validitet og gjennomførbarhet. I tillegg ble frekvensfordelingen oppgitt.

EVALUERING OG ETTERARBEID

Rangeringen av indikatorer ble oppsummert og sendt til ekspertgruppen. I rapporten er tilgang på data et avgjørende moment for valg av internasjonale indikatorer.

Ekspertgruppen delte pasientsikkerhetsindikatorerne inn i fire grupper på bakgrunn av skåringene i konsensusprosessen:

1. Umiddelbart tilgjengelig i alle europeiske helsesystem

2. Umiddelbart tilgjengelig i deler av de europeiske helsesystem
3. For øyeblikket ikke klar (gjennomførbar) for implementering i Europa. Anbefalinger om framtidig beslutning om implementering eller
4. Ikke gjennomførbar som pasientsikkerhetsindikator i Europa.

VURDERING AV SAIFA

Indikatorbeskrivelsen som ble utviklet i SimPatIE prosjektet dekker sentrale aspekt ved pasientsikkerhetsindikatorer. Det er nært beslektet med indikatorbeskrivelser fra USA, og er utprøvd ved beskrivelse av indikatorer som er benyttet i europeisk sammenheng. Beskrivelsen er klart teknisk og forutsetter god kjennskap til bruk og dokumentering av indikatorer fra tidligere.

En klar fordel med SAIFA er den relativt enkle skåringsmatrisen som er utviklet. På et ark finner en konsentrert definisjoner av dimensjoner som skal skåres og informasjon om skåringsdimensjon for de tre vurderingene som skal gjøres. Det er også det eneste skjema som avsatt egen plass til fri tekst. Det relativt enkle skåringskjema vil trolig øke sannsynligheten for at instrumentet blir tatt i bruk.

4. Appraisal of Indicators through research and Evaluation (AIRE)

Appraisal of Indicators through Research and Evaluation (AIRE) er et instrument utviklet i Nederland for vurdering av kvalitetsindikatorer. Det skiller seg dermed noe ut fra de andre instrumentene ved at det *ikke* primært er konstruert for valg av indikatorer. Det beskrives med to formål, henholdsvis å vurdere og å utvikle indikatorer. Ved å bruke vurderingsskjemaet for hvordan en indikator skal/kan beskrives har man en mal som kan brukes for utvikling av nye indikatorer. Det er ikke foreslått noen konkrete tiltak for hvordan en skal gjennomføre prosessene(14).

Felles med de andre instrumentene for konsensusprosesser er at de vurderer det vitenskapelige grunnlaget for indikatorer. AIRE vurderer i tillegg formål, relevans og organisasjonskontekst for utvikling av indikatorene, samt interessenters involvering i utvikling av indikatorer. Dette gjøres ved å rangere en rekke utsagn om den enkelte kvalitetsindikator innenfor fire kvalitetsområder.

FORARBEID

Ved vurdering av indikatorer forutsettes det en systematisk framstilling av informasjon om den enkelte indikator. Den skal beskrive fire ulike aspekt ved indikatoren slik at hvert enkelt punkt kan vurderes. Aspekter ved indikatorer som skal beskrives er:

1. Formål, relevans og organisasjonskontekst
 - a. formål til indikatoren er klart beskrevet
 - b. kriterier for temaområde er detaljert beskrevet
 - c. organisasjonskontekst, hvor skal den brukes
 - d. kvalitetsområde som indikatoren skal si noe om
 - e. helsetjenesteprosesser som indikatoren dekker
2. Interessenters involvering
 - a. bredt sammensatt gruppa av relevante profesjoner
 - b. involvering av alle i utviklingsfasen
 - c. formell oppslutning om indikatoren
3. Vitenskaplig grunnlag

- a. systematiske kunnskapsoppsummeringer benyttet for å beskrive grunnlaget
- b. indikatorens vitenskaplige grunnlag er enten evidensbasert eller basert på publiserte studier i vitenskaplige tidsskrift
- c. kritisk vurdering av evidensgrunnlaget
4. Ytterligere dokumentasjon, formulering og nytte
 - a. klart definert teller og nevner
 - b. klar populasjonsdefinisjon
 - c. strategi for risikojusteringer er vurdert og beskrevet
 - d. validitet
 - e. reliabilitet
 - f. indikatoren har tilstrekkelig diskrimineringssevne
 - g. pilottestet
 - h. omkostninger ved datainnsamling er vurdert
 - i. klare instruksjoner for presentering og fortolkning av indikatorresultat

Identifisering og kategorisering av de ulike aspektene gir en systematisk oversikt over indikatorer og hvordan de har blitt utviklet. Punktet med dokumentasjon av interessenters deltakelse skiller AIRE instrumentet fra andre systematiske presentasjoner av indikatorer. Sammenstillingen av informasjon som skal vurderes, fungerer som også som en oversikt over de ulike aspektene som skal dekkes ved utvikling av indikatorer.

RANGERINGER

For å rangere indikatorene er det utviklet et utsagn for hvert underpunkt i indikatorbeskrivelsen. Utsagnene er omformuleringer av kriteriene for indikatorenes egenkap som ”Indikatorens formål er klart beskrevet”. Utsagnene er koblet til utdypende definisjoner. Hvert utsagn rangeres på en skala fra 1 til 4 hvor man sier seg helt enig, delvis enig/uenig eller helt uenig i utsagnet.

For å beregne rangeringen til den enkelte indikator rangeres hvert hovedområde ved hjelp av summative skåringer og standardiseringer. Først summeres skåringene innenfor alle kriteriene i et område. Deretter regner man ut hvor stor prosentandel skåringen oppnår i forhold til mulig toppskår:

$$[(\sum X - \sum X_{\min}) / (\sum X_{\max} - \sum X_{\min})] * 100\%$$

$\sum X_{\max}$ er maksimumsskåren en indikator kan få om alle gir høyeste skåre til alle kriteriene, $\sum X_{\min}$ er minimumsskåren en indikator kan få om alle gir den laveste skåre til alle kriteriene.

Resultatene skal presenteres for det enkelte område, og ikke som en samlet skår for den enkelte indikator.

Det skal også gi en helhetsvurdering om indikatoren anbefales til videre bruk og det anbefales at indikatorene blir vurdert av minst fire eksperter/appraisers.

EVALUERING OG ETTERARBEID

Instrumentet er under utvikling, og det er så vidt vi kjenner til ikke publisert noen valideringsstudier som angir hvordan man kan/bør sammenfatte informasjonen om den enkelte indikator.

Informasjon og invitasjon til deltakelse i et internasjonalt valideringsarbeid for instrumentet er presentert på internett: <http://www.aire-instrument.com/>.

VURDERING AV AIRE

AIRE-skjemaet for indikatorbeskrivelse er omfattende, og samtidig er det klart formulert. I alt er det 20 områder som skal dokumenteres for den enkelte indikator. Det har også god beskrivelse av hvordan rangering av indikatorene skal gjøres. Det har ikke noen beskrivelse av hvordan konsensusprosesser bør gjennomføres, men er ved sitt språklig klart formulerte verktøy et godt instrument for indikatorbeskrivelse.

Sammenfatning

Formaliserte konsensusprosesser er i denne rapporten beskrevet både som prosess og verktøy. Arbeidsgruppen valgte å presentere Rand/UCLA-metodikken detaljert for å gi et innblikk i hvordan arbeidsprosessene er planlagt og bygd opp, og SAIRE og AIRE-skjemaene er presentert som eksempler på verktøy som benyttes i europeisk sammenheng.

BESKRIVELSE OG VURDERING

Et fellestrekk ved gjennomføring av konsensusprosesser er sammenfatting av kunnskapsgrunnlag etter et fastlagt mønster. Dette er etterfulgt av profesjonell gjennomgang og evaluering av en ekspertgruppe. Ekspertgruppens vurderinger er avgjørende for valg av indikatorer. Det er noen nyanser i hvordan kunnskapsgrunnlaget skal sammenfattes, men i hovedsak er det noen felles dimensjoner som ligger til grunn for valg av indikatorer.

En dimensjon gjelder indikatorens relasjon til helsetilstanden, hvorvidt indikatoren retter seg mot områder hvor det er forskjeller mellom faktiske og potensielle nivå for helse. Generelt vil en velge indikatorer som fanger opp et fenomen som opptrer relativt hyppig og gjelder mange, som for eksempel behandling av diabetes. Innenfor pasientsikkerhet er nettopp kravet om hyppighet eller frekvens for en indikator mindre rigid. Der kan en indikator som forekommer svært sjelden gi viktig informasjon om alvorlige brudd på pasientsikkerhet(1,13). Kravet om en klar populasjon, og hvem sin helsetilstand indikatoren representerer, er avgjørende for om en indikator velges. Alle de tre metodene/verktøyene i dette notatet har lagt vekt på denne dimensjonen som viser til indikatoren relevans og egnethet til å fange opp viktige hel-seaspekt.

En annen dimensjon er hvor egnet indikatoren er til å registrere påvirkning fra hel-sevesenet. Vil tiltak iverksatt av helsetjenesten føre til endring som blir registrert av

indikatoren, eller kan ikke effekten av enkelttiltak måles? En må stille spørsmål om indikatoren fanger opp alle tilfellene som skal registreres, og samtidig ikke registrere de tilfellene som ikke skal regnes med? Indikatorens validitet og reliabilitet er viktige når man skal velge indikatorer.

Den siste dimensjonen som går igjen i de ulike konsensusprosessene er tilgang på data og rapporteringsbyrde. For det første er det viktig å få oversikt om data er tilgjengelig på nasjonalt, såvel som på internasjonalt nivå. Deretter må man vurdere hvor stor rapporteringsbyrden er, om verdien av informasjon som er innbygd i en indikator, er større enn omkostnadene ved datainnsamling og rapportering.

De nevnte hoveddimensjonene kan skjematisk framstilles ved hjelp av det verktøyet som ble utviklet i forbindelse med SiMPatIE-prosjektet. Vi har her gjort et forsøk på å oversette verktøyet til norsk (Figur 5.1). Dette skåringskjemaet kan sees på som en kombinasjon av RAND/UCLA-metodikk og AIRE-instrumentet. Det har færre dimensjoner enn AIRE -instrumentet, men har ulike aspekt ved de tre hoveddimensjonene. Skalaen for skåring er lik den som er benyttet i RAND/UCLA-metodikken.

Figur 5.1: Skåringsverktøy for valg av indikatorer.

Dimensjon	Definisjon	Skåring 1-9
Relevant og passende	Indikator sentral (frekvens og alvorlighetsgrad)	
	innen et gitt område (populasjon/organisasjon)	
Valide og reliable	Konstruksjonsvaliditet	
	Intern konsistens	
	Sensitivitet/spesifitet	
	Reliabilitet	
Gjennomførbar	Vurder tilgang på eventuelle data, samt byrde ved eventuelt merarbeid for helsepersonell ved innsamling av data	

Alternativt kan en mer detaljert inndeling gjøres slik det er antydnet ved de stiplede linjene i figuren. Krav til detaljeringsnivå bør vurderes i det enkelte tilfelle. Ulempen med dette verktøyet er at en ikke vurderer opplysninger om kontekst eller formål med indikatoren. De opplysningene er implisitte for de rammene verktøyet er utarbeidet for. Indikatorene skal si noe om pasientsikkerhet i sykehus. Skal instrumentet benyttes til andre, mer generelle formål som vurdering av kvalitet bør de kategoriene innarbeides i skjemaet.

AIRE- Instrumentet har i tillegg en oppsummerende kategori for anbefalt bruk av indikatoren. Oppsummering av disse skåringene kan gjøres uavhengig av resten av instrumentet. En analyse av instrumentet for å se på sammenhengen mellom overordnet anbefaling om bruk av indikator i forhold til skåringene på kriteriene, vil gi viktig informasjon om hvordan ekspertgrupper vektlegger ulike egenskaper til indikatorer.

Ulik vektlegging av kriterier kan være systematisk skjevfordelt mellom ulike grupper.

MOMENTER TIL DISKUSJON

Det er flere forhold som vil være med å påvirke utfallet av formaliserte konsensusprosesser(6,8-10,13,18). Det er sannsynlig at holdninger til oppdraget i utgangspunktet vil være med på å påvirke utfallet av konsensusprosesser. Mandatet, fokuset og holdningen til oppgaven som skal utredes vil også kunne ha innflytelse på resultatet. Dette stiller krav til planleggingsfasen og involvering av ulike interessenter i prosjektet.

Et annet aspekt er sammensetningen av ekspertgruppen og rammebetingelser de arbeider under. Ved sammensetningen av ekspertgruppen er det anbefalt at faglig ekspertise og uavhengighet skal vektlegges. Hvilke kriterier man velger å legge til grunn for utvelgelsen kan ha innflytelse på utfallet. Erfaringer fra Australia antyder sammensetningen av interessegrupper som skal rangere, kunnskapsgrunnlaget, og egenskaper ved paneldeltakerne også hadde en effekt på rangering av indikatorene(11). Det bør være representanter fra de ulike interessegrupper som blir berørt, både de som ønsker indikatorer, de som skal levere indikatorer og brukerrepresentanter. Når man vurderer resultatene fra et ekspertpanel må en derfor om mulig ta hensyn til hvordan panelet er sammensatt med tanke på spesialisitet(8). Størrelsen på gruppen kan også ha innflytelse på resultatet, men det er lite forskning på området, noen studier viser at det ikke bør være færre enn 6, og helst ikke flere enn 12 deltakere(11). Både arbeidsbetingelsene og sammensetningen av ekspertgruppene kan ha direkte betydning for utfallet, det vil si hvilken konsensus, som oppnås (11,19).

Presentasjon av kunnskapsgrunnlag og evalueringsverktøy bør være konsistent over tid. Det er sentralt for å sikre at resultatene fra en konsensusprosess kan reproduseres. Tanken bak er at gitt samme presentasjon av informasjon og bruk av like prosesser vil resultatene bli tilnærmet like.

Gjennomføring og styring av konsensusmøtene bør planlegges godt. For eksempel viser amerikanske erfaringer at fredag til lørdag har vært gode valg for ekspertsamlinger. I norsk sammenheng vil trolig samling på vanlige yrkesdager være en forutsetning for å samle ekspertgruppen. Vi har små eller ingen erfaringer fra norsk kontekst med gjennomføring av konsensusgrupper og bør samle egne erfaringer på dette området.

Sammenfatting av resultater fra formaliserte konsensusprosesser bør også følge noenlunde samme mønster. Overgang fra papirbaserte spørreskjema til nettbaserte spørreskjema vil trolig både forenkle og effektivisere bruken av formaliserte konsensusprosesser(20).

ARBEIDSGRUPPENS ANBEFALINGER

Metodikken forutsetter et profesjonelt forarbeid av en tilrettelegger som administrerer og dokumenterer prosessene. Ulempen er at det er relativt tid- og ressurskrevende å gjennomføre formaliserte konsensusprosesser. Fordelen er at det ligger et systematisk og grundig arbeid til grunn for indikatorer som skal gi uttrykk for pasientsikkerhet.

Arbeidsgruppen anbefaler at man ved utvikling av indikatorer innenfor et område tar utgangspunkt i eksisterende indikatorer og beskrivelsen av dem. Ved behov for utvikling av nye indikatorer bør de beskrives på tilsvarende måte. Når man skal velge indikator anbefaler vi å følge en modifisert RAND/UCLA-metodikk. Ved å benytte ekspertpanel som har en viss størrelse og en klart dokumentert metode for gjennomføring av konsensusmøter oppnår en større åpenhet og kunnskap om hvorfor noen indikatorer blir valgt og andre ikke. Dokumentasjon av disse prosessene er viktig i utviklingen og forståelsen av hvordan indikatorene kan bidra til en bedre helsetjeneste.

Aire-instrumentet har den klareste formuleringen av hvordan indikatorene skal beskrives. Den framstår i første omgang som noe mer omfattende enn SAIFAs indikatorbeskrivelse, men en sammenstilling av de to skjemaene viser til dels overlappende kategorier. Vi anbefaler at eventuelle indikatorbeskrivelser legges tett opp mot den man finner i AIRE-instrumentet. Simpatieprosjektet sitt skåringsverktøy dekker de sentrale hoveddimensjonene indikatorene er beskrevet i forhold til. Det har potensial til å ta vare på nyanser i beskrivelsen av ulike indikatorer, samtidig som det ikke gir et så detaljert bilde at det blir ressursskrevende å analysere og fortolke. Vi anbefaler at man dokumenterer og synliggjør de valg som gjøres ved utvikling og valg av kvalitetsindikatorer. I denne rapporten har vi beskrevet nærmere noen ulike konsensusprosesser og mulige gjennomføringer av dem. Ved å benytte seg av internasjonalt anerkjent prosedyrer og normer for gjennomføring av prosessen sikrer man en enhetlig og god dokumentasjon av indikatorer som skal benyttes innenfor pasientsikkerhet.

På grunn av de relativt store omkostningene vil full implementering av metodikken trolig være best egnet ved utvikling og valg av indikatorer på overordnet nivå, nasjonalt eller internasjonalt. Arbeidsgruppen anbefaler likevel at all utvikling av indikatorer for pasientsikkerhet dokumenteres etter anbefalinger og metoder som er beskrevet i denne rapporten.

Litteraturliste

Reference List

1. Vincent C. Patient Safety. Edinburgh: Elsevier Churchill Livingstone, 2006.
2. Institute of Medicine. Envisioning the national health care quality report. Washington, D.C.: National Academy Press, 2001.
3. Institute of Medicine. Patient Safety: Achieving a New Standard for Care. 2003.
4. Walshe K, Boaden RJ. Patient safety: research into practice. Maidenhead: Open University Press, 2006.
5. Rygh LH, Saunes IS. Utvikling og bruk av kvalitetsindikatorer for spesialisthelsetjenesten. nr 6-2008 ed. Oslo: Nasjonalt kunnskapssenter for helsetjenesten, 2008.
6. Campbell SM, Braspenning J, Hutchinson A et al. Research methods used in developing and applying quality indicators in primary care. Qual Saf Health Care 2002; 11: 358-64.
7. Campbell SM, Braspenning J, Hutchinson A et al. Research methods used in developing and applying quality indicators in primary care. BMJ 2003; 326: 816-9.
8. Fitch K, Bernstein S, Aguilar MS, Burnand B, LaCalle JR, Lazaro P et al. The Rand/UCLA Appropriateness Method User's Manual. Santa Monica, CA: RAND, 2001.
9. Black N, Murphy M, Lamping D et al. Consensus development methods: a review of best practice in creating clinical guidelines. J Health Serv Res Policy 1999; 4: 236-48.
10. Jones J, Hunter D. Consensus methods for medical and health services research. BMJ 1995; 311: 376-80.
11. Halcomb E, Davidson P, Hardaker L. Using the consensus development conference method in healthcare research. Nurse Res 2008; 16: 56-71.
12. National Institute of Health. NIH Consensus Development Programme. 2009.

13. Kristensen S MJBP. Establishing a Set of Patient Safety Indicators. Safety Improvements for Patients in Europe Working package 4. Denmark: The ESQH-office for Quality Indicators, 2007.
14. de Koning Dr.Johan S. Appraisal of Indicators through Research and Evaluation. National Institute for Public Health and the Environment, 2009.
15. Kelly E, Hurst J. Health Care Quality Indicators Project Initial Indicators Report. OECD HEALTH WORKING PAPERS 22. Paris: OECD, 2006.
16. Brook RH, Chassin MR, Fink A et al. A method for the detailed assessment of the appropriateness of medical technologies. *Int J Technol Assess Health Care* 1986; 2: 53-63.
17. Romano PS. Selecting indicators for Patient Safety at the health system level in OECD countries. Summary of recent US experience. 2008.
18. Millar J, Mattke S, and the Members of the OECD Patient Safety Panel. Selection Indicators for Patient Safety at the Health Systems Level in the OECD Countries. 18. Paris: OECD, 2004.
19. Hermann RC, Palmer RH. Common ground: a framework for selecting core quality measures for mental health and substance abuse care. *Psychiatr Serv* 2002; 53: 281-7.
20. Deshpande AM, Shiffman RN, Nadkarni PM. Metadata-driven Delphi rating on the Internet. *Comput Methods Programs Biomed* 2005; 77: 49-56.