



Evaluering av pilot for systematisk innhenting av ekspertvurderinger

oslo**economics**



UNIVERSITETET
I OSLO

Tittel: Evaluering av pilot for systematisk innhenting av ekspertvurderinger

Utarbeidet av: Oslo Economics

Oppdragsgiver: CONNECT (Norwegian Cancer Precision Medicine Implementation Consortium)

Publisert: Oktober 2025

Rapportnummer: 2025-77

Kontaktperson: Erik Magnus Sæther / Partner

E-post: ems@osloeconomics.no

Tel: +47 940 58 192

Foto/illustrasjon forside: Microsoft Stock Images

Innhold

1. Systematisk innhenting av ekspertvurderinger	1
1.1 Ekspertvurderinger blir en stadig viktigere kilde til informasjon	1
1.2 Pilotprosjekt i Norge	1
2. Rammeverk for innhenting av strukturerde klinikerinnspill	3
2.1 Økonomisk evaluering av nye legemidler	3
2.2 Rammeverk for systematisk innhenting av ekspertvurderinger	4
3. Praktisk gjennomføring av piloten	6
3.1 Forberedelser	6
3.2 Individuelle intervjuer	7
3.3 Aggregering	7
3.4 Modellering og analyse	7
4. Vurdering av den praktiske gjennomføringen	8
4.1 Det individuelle intervjuet	8
4.2 Konsensumøtet	8
4.3 Bruk av metoden i praksis	9
5. Drøfting	10
5.1 Antall eksperter og rekruttering	10
5.2 Forberedelse av eksperter	10
5.3 Gjennomføring av konsensumøte	11
5.4 Fokus for videre forskning	12
6. Referanser	13

1. Systematisk innhenting av ekspertvurderinger

Innspill fra kliniske eksperter blir i økende grad brukt som informasjonskilde i evalueringer av om nye behandlinger skal finansieres av den offentlige helse- og omsorgstjenesten. De siste årene er det utviklet flere metoder for systematisk innhenting av innspill fra eksperter. Slike metoder kan være nyttig når kliniske data er begrenset, men beslutninger om offentlig finansiering av nye metoder likevel må tas.

Metoder for systematisk innhenting av ekspertvurderinger gjør det mulig å kvantifisere usikkerhet, og dermed styrke beslutningsgrunnlaget i økonomiske analyser av nye metoder. Flere rammeverk er utviklet de siste årene, men det er begrenset med erfaringer fra gjennomføringen av slik datainnhenting i praksis. Oslo Economics, i samarbeid med Universitetet i Oslo og på oppdrag fra CONNECT, har derfor gjennomført flere pilotprosjekter for å evaluere rammeverket i en norsk kontekst. Formålet med arbeidet har vært å bidra til kunnskap om den praktiske gjennomføringen av systematisk innhenting av ekspertvurderinger.

1.1 Ekspertvurderinger blir en stadig viktigere kilde til informasjon

Eksperters vurderinger er stadig mer sentrale for å forstå pasientgrupper, behandlingseffekt og kostnader, i tillegg til usikkerhet knyttet til disse forholdene. Det er derfor behov for metoder for å innhente klinikers vurderinger på en strukturert måte, som også tar hensyn til usikkerheten i vurderingene og bidrar til å redusere skjevhet i tilbakemeldingene.

National Institute for Health and Care Excellence (NICE) skiller i sin manual for evaluering av nye tiltak i helsetjenesten mellom det de kaller «expert opinion» og «structured expert elicitation», hvorav det siste viser til metoden som piloteres i dette arbeidet (Oakley, et al., 2025). Førstnevnte handler i større grad om at man bruker eksperter til å supplere, validere og tolke data fra randomiserte studier. Systematisk innhenting av ekspertvurderinger innebærer bruk av en protokoll, der spørsmål er klart definert på forhånd og utfallet av

prosessen er en sannsynlighetsfordeling for å representere usikkerhet hos ekspertene. NICE manualen trekker frem at en styrke ved systematisk innhenting av vurderinger er at det bidrar til å redusere skjevhet, samtidig som det tar hensyn til usikkerhet rundt vurderingene.

Et eksempel på bruk av metoden er nederlandske myndigheter (Zorginstituut Nederland), som anvender strukturerte ekspertvurderinger som supplement til annen dokumentasjon for viktige parametere (f.eks. langtidseffekt) som er krevende å estimere direkte fra studier. Innhenting av informasjon fra eksperter gjennomføres etter etablerte rammeverk (nærmere beskrevet i kapittel 2.2). Eksperter kan rekrutteres fra klinisk praksis, forskningsmiljø og noen ganger profesjonsforeninger (f.eks. Legeforeningen) eller pasient- og brukerorganisasjoner. Prosessen er strukturert: ekspertene får samme informasjon, spørsmålene standardiseres, og usikkerheten kvantifiseres (med sannsynlighetsfordelinger).

1.2 Pilotprosjekt i Norge

1.2.1 Pilotering av rammeverk

Systematisk innhenting av ekspertvurderinger i helseøkonomiske analyser har blitt brukt i flere land, eksempelvis Storbritannia, Nederland, Canada og Australia. Rammeverk som SHELF og IDEA gir strukturerte prosedyrer, men det er fortsatt få eksempler på omfattende, systematisk bruk i beslutningsprosesser på nasjonalt nivå.

For å utvikle metoden og eventuelt benytte informasjonen fra denne typen datainnhenting ved innføring av nye legemidler er det behov for praktiske erfaringer. Det er nødvendig å teste hvordan rammeverket fungerer i praksis, og kunnskap om hvordan metoden fungerer, hva som er viktig å hensynta og når den bør benyttes før den eventuelt implementeres bredt. Dette gjelder blant annet i norsk kontekst.

Pilotstudier kan avklare om prosedyrene er forståelige for ekspertene, om tidsbruk og ressursbruk er realistisk, om usikkerhet kan kvantifiseres på en konsistent måte, hvordan resultatene best integreres i helseøkonomiske modeller og hvilken betydning informasjonen kan ha for beslutninger. I tillegg gir pilotstudier mulighet til å identifisere metodologiske utfordringer, som gruppepress, skjevheter i vurderinger eller vansker med å formulere spørsmål på en måte som gir meningsfulle anslag.

1.2.2 Pilot finansiert av CONNECT

CONNECT er et offentlig-privat konsortium som jobber for å innføre presisjonsmedisin innen kreftbehandling raskere. Partene i CONNECT prøver i fellesskap ut nye løsninger innen presisjonsmedisin for å gjøre praksis på området bedre. En arbeidsgruppe i CONNECT har gitt Oslo Economics og Universitetet i Oslo i oppdrag å gjennomføre pilotstudier der rammeverket for systematisk innhenting av ekspertvurderinger testes på konkrete saker i Norge. Pilotprosjektet som presenteres i denne rapporten ble påbegynt høsten 2023 og ferdigstilt høsten 2025. Denne evalueringsrapporten dokumenterer erfaringene fra dette arbeidet.

Oslo Economics har i samarbeid med Universitetet i Oslo gjennomført to tidligere piloter av metoden for følgende indikasjoner:

- Selpercatinib til behandling av kreft i skjoldbruskkjertelen
- Selpercatinib til behandling av ikke-småcellet lungekreft

Med bakgrunn i de tidligere pilotene gjennomførte vi våren 2025 en tredje pilot. Hensikten var todelt; vi ønsket for det første teste hvordan rammeverket fungerte i praksis, samt innhente erfaringer med rammeverk og metode i etterkant. Denne rapporten gjengir disse erfaringene.

Denne piloten bygger i hovedsak på rammeverk gitt av SHELFL, basert på Bojke et al (2021) (se kapittel 3 for en fullstendig beskrivelse av den praktiske gjennomføringen).

Evaluering av datainnhenting

I etterkant av selve piloteringen utformet vi et spørreskjema, med hensikt om å innhente erfaringer fra ekspertene som hadde deltok i piloten. Skjemaet ble utformet i Questback og ble distribuert via e-post til ekspertene som deltok i den tredje piloten. Vi mottok svar fra fem av syv eksperter.

Tema for spørreskjema var følgende

- Det individuelle intervjuet (samtalet, informasjonspakken i forkant, teknisk gjennomføring og verktøyet som benyttes)
- Konsensumøtet (felles diskusjon, fagsammensetning, møteledelse)

Hvorfor systematisk innhenting av ekspertvurderinger?

Systematisk innhenting av ekspertvurdering benyttes for å estimere informasjon (for eksempel om effekt av behandling) ved svak eller fraværende evidens. Anta at det finnes en ny behandling, men evidens knyttet til effekt (total overlevelse (OS) og progresjonsfri overlevelse (PFS)) kun er dokumentert i en fase 2 studie (studie uten kontrollgruppe) med kort oppfølgingsperiode. Basert på informasjonen fra studien er det ikke mulig å etablere relativ effekt ettersom studien ikke sier noe om hvordan det går med pasienter som ikke får behandlingen. I slike tilfeller kan systematisk innhenting av ekspertvurderinger gi:

En overordnet vurdering av om resultatene fra den kliniske studien kan overføres til lokale forhold – Basert på en sammenligning av studiepopulasjonen med lokale forhold.

Et estimat på behandlingseffekten for kontrollarmen som grunnlag for å etablere estimat for relativ effekt - Basert på lokal registerdata (hvis tilgjengelig) og klinisk erfaring med pasientgruppen.

Et estimat på de langsiktige effektene av behandlingen - Anslag for hvordan det går med pasientene over tid, basert på resultater fra den kliniske studien, lokale data og klinisk erfaring.

- Bruk av metoden i praksis (når er det verdt arbeidet, rekruttering av eksperter, geografisk variasjon, kilder til usikkerhet)

Evalueringsrapporten

Denne rapporten bygger videre på innspill gitt både i individuelle intervjuer og konsensumøte. Oakley et al (2025) ved The Decision Support Unit External Assessment Centre¹ (DSU) har utarbeidet en omfattende rapport der forslag til beste praksis ved bruk av systematiske ekspertvurderinger presenteres, basert på tidligere bruk i beslutningstaking og litteratur. Arbeidet er finansiert av NICE. I tillegg har en ISPOR task force oppsummert styrker og svakheter ved fem ulike protokoller for strukturert innhenting av klinikerinnspill, og basert på dette kommet med anbefalinger om bruk (Soares, et al., 2024). Erfaringer fra den praktiske gjennomføringen av vår pilot sees i lys av tidligere arbeid.

¹ DSU er et forskningssenter som gjør oppdragsbasert forskning på vegne av NICE for å støtte deres arbeid.

2. Rammeverk for strukturert datainnhenting

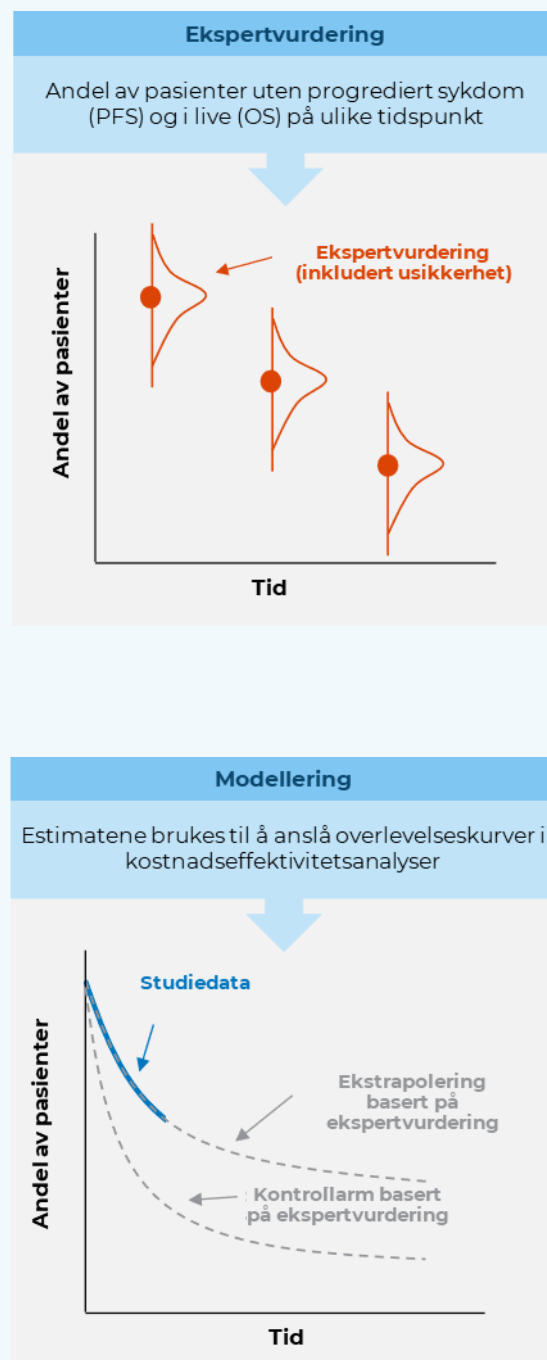
Før et nytt legemiddel finansieres av den offentlige helsetjenesten vurderes det om legemiddelet oppfyller de tre prioriteringskriteriene; nytte, ressursbruk og alvorlighet. Dette gjøres for å sikre at legemiddelet som innføres gir større verdi enn å bruke ressursene på andre tiltak. Samfunnsøkonomiske beslutninger tas under usikkerhet, og denne øker når tilgjengelig dokumentasjon er begrenset. Dette gjelder særlig for avanserte terapier og behandlinger rettet mot små pasientpopulasjoner. Systematisk innhenting av ekspertvurderinger kan benyttes for å etablere et mer robust beslutningsgrunnlag der ordinære data mangler.

2.1 Økonomisk evaluering av nye legemidler

Før et nytt legemiddel finansieres av den offentlige helsetjenesten gjøres det vurderinger av ulike forhold. Før innføring i Norge vurderes legemidler av europeiske legemiddelmyndigheter (EMA). I denne evalueringen vurderer EMA om legemiddelet virker (har positiv nytte for pasienten) og om det er trygt å ta i bruk. Deretter vurderer norske myndigheter hvordan innføringen påvirker ressursbruken i helsetjenesten i Norge. Basert på informasjon fra produsenten gjennomfører Direktoratet for medisinske produkter (DMP) en metodevurdering, som danner grunnlag for beslutningen om offentlig finansiering. Den nasjonale vurderingene bygger på de tre prioriteringskriteriene som gjelder i den norske helse- og omsorgstjenesten (Meld. St. 21, 2024–2025): **nytte, ressursbruk og alvorlighet**.

- **Nytte** handler om hvor stor helsegevinst legemidlet gir, for eksempel forlenget levetid, bedre livskvalitet eller færre symptomer.
- **Ressursbruk** dreier seg om kostnadene ved legemidlet. Jo høyere kostnaden er, jo lavere prioritet vil legemiddelet få.
- **Alvorlighet** betyr at sykdommens alvorlighetsgrad spiller en rolle – jo mer alvorlig sykdom, desto høyere er betalingsvilligheten for legemidler for sykdommen.

Figur 1 Eksempel på hvordan systematisk innhenting av ekspertvurderinger kan brukes for å informere kontrollarm og langtidsoverlevelse i en helseøkonomisk analyse



Illustrasjon: Oslo Economics

Sammen skal denne vurderingen sikre at nye legemidler som tas i bruk, gir verdi både for pasientene og for helsetjenesten som helhet, sammenlignet med verdien av å bruke ressursene på andre tiltak. For å vurdere om disse kriteriene er oppfylt må det gjennomføres økonomiske analyser som sammenligner helsegevinster og kostnader, sett opp mot sykdommens alvorlighet.

2.1.1 Beslutninger under usikkerhet

Beslutning om offentlig finansiering av et legemiddel må tas under usikkerhet, både om behandlingens kostnader og helseeffekter. Fordi en rekke kostnads- og nyttevirksomheter skal belyses må analysene bygge på flere informasjonskilder, ikke bare informasjon fra kliniske studier.

Flere nye legemidler som er utviklet de siste årene har potensielt stor behandlingseffekt, men også høye kostnader. Det er også større usikkerhet knyttet til behandlingseffekten til nye legemidler i dag enn hva det har vært tidligere. Dette skyldes blant annet at flere behandlinger er rettet mot små pasientgrupper, der det er mer krevende å gjennomføre store randomiserte studier (fase 3-studier, det vil si studier som sammenligner effekten av behandlingen mot en kontrollgruppe). I noen tilfeller er det heller ikke mulig å gjennomføre denne typen studier, for eksempel av praktiske eller etiske årsaker. Dette gjør at beslutningstakere har begrenset med informasjon når de skal fatte en beslutning om behandlingen skal innføres og finansieres av den offentlige helsetjenesten. I enkelte tilfeller vil beslutningstaker kunne utsette beslutningen og vente på mer informasjon, mens i andre tilfeller vil det ikke komme mer informasjon for å belyse saken. Beslutningstaker må gjøre vurderinger knyttet til behovet og verdien av å innhente mer informasjon. Fra samfunnets side er det likevel ønskelig at det fattes en beslutning om metoden skal innføres, ettersom det å vente også har en kostnad i form av tapt helse for pasienter som ikke får tilgang til behandling.

De siste årene er det gjort en rekke krevende vurderinger knyttet til innføring av nye behandlingstilbud for små pasientgrupper. I mange saker der behandlingen ikke innføres er begrunnelsen fra beslutningstakere at kostnaden ved behandlingen er for høy sett i forhold til usikkerheten knyttet til gevinstene av behandlingene.

2.1.2 Beslutninger basert på flere kilder

Når det fattes beslutninger om offentlig finansiering av nye legemidler i Norge, er dokumentasjonsgrunnlaget avgjørende. Innføring av et nytt legemiddel håndteres ofte som en

irreversibel beslutning, da det er krevende å fjerne et behandlingstilbud som allerede er innført.

Randomiserte kontrollerte studier (RCT-er) regnes som gullstandarden ved vurdering av tiltak, fordi de på en god måte kan dokumentere både effekt og sikkerhet av ny behandling. Resultater fra slike studier er i mange tilfeller et sentralt informasjonsgrunnlag for de matematiske simuleringsmodellene som brukes til å beregne kostnader og nytte av nye legemidler. Data fra systematiske oversikter, metaanalyser, real-world evidence (registerdata og erfaring fra klinisk praksis) brukes også i helseøkonomiske analyser, enten alene, i fravær av data fra RCT, eller som et tillegg til RCT.

2.2 Rammeverk for systematisk innhenting av ekspertvurderinger

Det finnes flere etablerte protokoller eller rammeverk for systematisk innhenting av ekspertvurderinger; Cooke's classical method, Delphi methods, IDEA, MRC reference protokoll og SHELF.

De ulike rammeverkene har mange fellestrekk, men også flere sentrale forskjeller

- Cooke's classical method vektlegger for eksempel individuell kalibrering og gir størst vekt til eksperter som viser seg mest treffsikre (Cooke, 1991)
- Delphi-metodene bygger på flere runder med anonyme vurderinger for å oppnå konsensus (Dalkey & Helmer, 1963)
- IDEA kombinerer individuell estimering og gruppegjennomgang før man samler resultatene (Hemming, et al., 2018)
- MRC reference protocol er en mer generell veileder for hvordan datainnhenting bør planlegges og dokumenteres (Bojke, et al., 2021)
- SHELF gir en praktisk prosedyre for gruppebaserte vurderinger med eksplisitt kvantifisering av usikkerhet og felles fordelinger (Oakley & A, u.d.)

En relevant forskjell mellom metodene er at én metode (Cooke) vektlegger kalibrering, mens andre (Delphi, IDEA, SHELF) fokuserer på gruppedynamikk og konsensus. Felles er ønsket om å redusere skjevhet og gjøre ekspertvurderinger etterprøvbare og nyttige i beslutningsprosesser.

En sentral forskjell mellom protokollene er om man benytter seg av både individuelle og

gruppediskusjoner for å etablere anslag. Noen protokoller inkluderer både innhenting av anslag gjennom individuelle anslag, samt et konsensusmøte der formålet er at ekspertene skal enes om et anslag. Andre protokoller benytter matematisk aggregering av anslag gitt i individuelle intervjuer.

2.2.1 Eksempler på bruk av rammeverket i helseøkonomiske analyser

Systematisk innhenting av ekspertvurderinger kan brukes på ulike måter. Pilotene som er gjennomført har undersøkt hvordan rammeverket kan brukes til å informere en kontrollarm (i tilfeller der det ikke er gjort en fase 3-studie) eller langtidsvirkninger av behandlingen (effekt utover hva som er observert i de kliniske studiene). Dette kan for eksempel være ved å systematisk innhente anslag for progresjonsfri overlevelse (PFS) og total overlevelse (OS) fra kliniske eksperter. I tillegg bes ekspertene beskrive og kvantifisere usikkerheten i anslagene. På den måten kan man etablere et mer robust beslutningsgrunnlag der ordinære data mangler.

Metoden kan anvendes i flere situasjoner enn dette og er relevant i beslutningssituasjoner der tilgjengelig dokumentasjon er begrenset. Klinikere konsulteres allerede i dag som ledd av metodevurderinger og bruk av dette rammeverket kan øke nøyaktigheten i disse anslagene. Eksempler på slike parametere er alder på pasientene, prevalens, insidens, bruk av legemidler i norsk klinisk praksis, herunder gjennomsnittlig dosering og forventet varighet av behandling. Eksperter kan også gi anslag på markedsandeler til bruk i budsjettkonsekvensanalyser. Man må imidlertid vurdere ressursbruk opp mot verdi av økt nøyaktighet i en beslutningsprosess.

For sentrale parametere som effekt vil metoden særlig være relevant for avanserte terapier, sjeldne sykdommer eller små pasientpopulasjoner, der det ofte er lite informasjon i de kliniske studiene, samtidig som beslutningstakere må kunne vurdere effekt, overlevelse og kostnadseffektivitet på tilgjengelig kunnskapsgrunnlag.

3. Praktisk gjennomføring av piloten

Det er per i dag ikke fullstendig konsensus knyttet til hvordan datainnhenting bør gjennomføres ved systematisk innhenting av ekspertvurderinger. Det ble både gjennomført individuelle samtaler med hver enkelt ekspert, og et konsensusmøte der ekspertene ble samlet.

Den praktiske gjennomføringen av piloten omfattet følgende steg: Forberedelser, individuelle intervjuer, aggregering og modellering og analyse. I det videre beskriver vi hvordan hver del ble gjennomført i forbindelse med denne piloten.

3.1 Forberedelser

For piloten ble pemigatinib til behandling av kolangiokarsinom valgt som case. Dette valget ble gjort basert på innspill fra DMP og CONNECT. Pasientpopulasjonen ble avgrenset til:

- Pasienter med lokalt avansert, metastatisk og/eller inoperabel kolangiokarsinom (gallegangskreft)
- Gjennomgått minst én tidligere linje med systemisk behandling
- Fibroblastvekstfaktor-reseptor 2-rearrangering (FGFR2+)

Rekruttering av kliniske eksperter

Et sentralt spørsmål i piloten var hvor mange eksperter som skulle inkluderes. Bojke et al (2021) anbefaler minimum fem eksperter, men det finnes ingen streng begrunnelse for et nøyaktig antall. Det er viktig at det er nok eksperter for å sikre at alle ekspertiseområder er representert, og at de som skal benytte resultatene er trygge på at nok eksperter har gitt sine innspill.

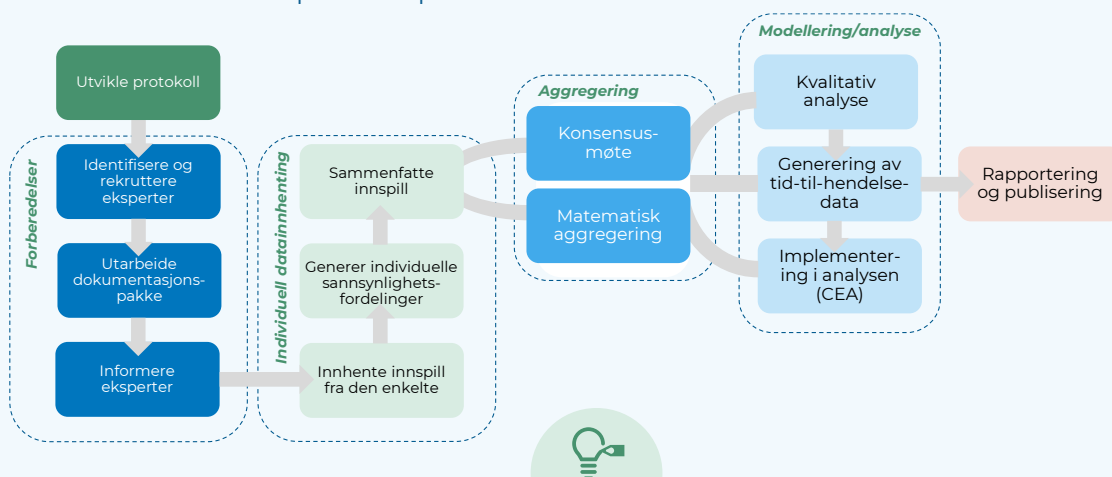
Vi har benyttet en «snøball»-metode for å identifisere relevante eksperter i Norge. Arbeidsgruppen i CONNECT, DMP og Norsk onkologisk foreningen ga i første omgang innspill til relevante eksperter. I henvendelsen til ekspertene ble både bakgrunn og formål med piloten beskrevet, samt hva dataen skulle brukes til. Ekspertene ble også bedt om å foreslå andre kandidater. Deltakerne ble rekruttert via epost og telefon.

Oppgaven i piloten

Utgangspunktet for oppgaven var den kliniske studien FIGHT-202 (Abou-Alfa, et al., 2020). Det var en enarmet fase 2 studie som rekrutterte pasienter som beskrevet over, og behandlet disse med pemigatinib frem til sykdomsprogresjon.

Oppgaven som ble gitt ekspertene var å gi anslag for progresjonsfri overlevelse (PFS) ved 6 og 12 måneder og totaloverlevelse (OS) ved 6, 12 og 36.

Figur 2: Oversikt over arbeidsprosess i piloten



Proessen og rammeverket ble utviklet basert på innspill fra medlemmer i arbeidsgruppen i CONNECT, inkludert representanter fra myndigheter, akademia og industri.

Anslagene skulle inkludere usikkerhet, og omfattet både en situasjonen der pasientene fikk og ikke fikk behandling med pemigatinib. Målet med oppgaven var å få anslag for PFS og OS for pasientpopulasjonen for en tenkt kontrollarm i FIGHT-202-studien og for intervensjonsarmen dersom behandling med pemigatinib hadde blitt innført i norsk klinisk praksis.

I forkant av intervjuet ble en informasjonspakke sendt til de rekrutterte ekspertene. Denne inneholdt en beskrivelse av pasientpopulasjoner og intervensjon, beskrivelse av metode for informasjonsinnhenting (Roulette metoden + eksempel, SHELF Expert briefing) og konkrete spørsmål for det individuelle intervjuet. DMP og representanter fra arbeidsgruppen i CONNECT var involvert i arbeidet med dokumentasjonspakken og ga innspill til forskergruppen i dette arbeidet.

3.2 Individuelle intervjuer

Det ble satt opp individuelle intervjuer med alle ekspertene. Disse ble gjennomført både fysisk og digitalt basert på ekspertens preferanser. I forkant av samtalen fikk eksperten informasjon om formålet med piloten, samt dokumentasjonspakken som var grunnlaget for samtalen. Dokumentasjonspakken ble igjen sendt ut et par dager før intervjuet som en påminnelse om nødvendige forberedelser. To ressurser fra Oslo Economics stilte i alle intervjuene, og det var samme person som ledet intervjuet med alle ekspertene.

3.3 Aggregering

I etterkant av de individuelle intervjuene inviterte vi til et fysisk konsensumøte. Fem av syv eksperter hadde mulighet til å delta. Tre av disse deltok fysisk, mens to deltok digitalt.

I konsensumøtet presenterte vi anslagene som hadde blitt gitt i de individuelle intervjuene (anonymt), og ga ekspertene mulighet til å enes om en endelig fordeling.

I forkant av møtet ble det utarbeidet et gjennomsnitt av alle fordelinger gitt i individuelle intervjuer. Gjennomsnittet ble i konsensumøtet presentert som et utgangspunkt for diskusjonen. Se faktaboks for illustrasjon av hvordan anslagene ble presentert. Etter fordelingen ble presentert, ble det samme verktøyet som ble benyttet i individuelle intervjuet åpnet, slik at ekspertene kunne justere anslaget og usikkerheten rundt anslaget.

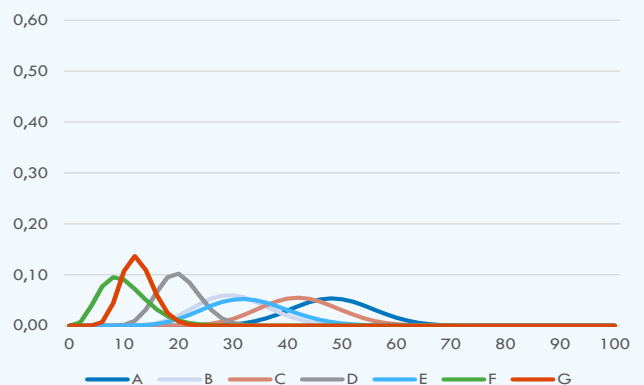
Underveis i konsensumøte ble anslagene ekspertene ble enige om fortløpende lagt inn i en

Presentasjon av individuelle anslag i konsensumøte

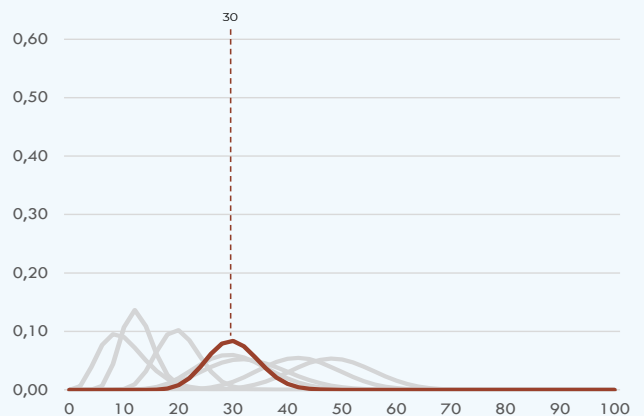
Under illustrerer vi hvordan klinikerne fikk presentert anslagene fra de individuelle intervjuene. De ble vist to grafer, der den første (øverste felt) viser alle fordelingene. Den andre (nederste felt) illustrerte et gjennomsnitt av anslagene, både foreslått median (som markert) og usikkerhet rundt denne medianen.

I tillegg ble klinikerne vist hvordan den røde fordelingen på nederste felt vil se ut i det tekniske verktøyet som benyttes, for å sørge for en tydelig forståelse av resultatet.

Alle anslag gitt i individuelle intervjuer



Gjennomsnitt av alle anslag



analyse for å utarbeide Kaplan-Meier plot over total og progresjonsfri overlevelse for de ulike populasjonene. Disse analysene ble vist og diskutert avslutningsvis i møtet.

3.4 Modellering og analyse

I etterkant av datainnhenting ble det gjennomført analyser. Denne delen av piloten er ikke tema for denne rapporten. Analyse av datamaterialet fra denne piloten er fortsatt under arbeid, med mål om å publisere resultater i vitenskapelige tidsskrift når de foreligger.

4. Vurdering av den praktiske gjennomføringen

Flertallet av ekspertene som deltok ga tilbakemeldinger om at de opplevde øvelsen som nyttig og at det tekniske rammeverket og gjennomføringen fungerte godt. Det trekkes imidlertid frem at gode informasjonspakker og tid til forberedelse er viktig for at datainnhenting skal resultere i et godt informasjonsgrunnlag som er relevant for beslutninger om innføring av nye metoder.

Ekspertene ga innspill til gjennomføringen og metoden både i sine individuelle intervjuer, og i spørreundersøkelsen som ble distribuert i etterkant av konsensusmøtet. I det følgende presenterer vi tilbakemeldingene fra ekspertene som deltok i piloten.

4.1 Det individuelle intervjuet

Det individuelle intervjuet var det første møtet med rammeverket for fem av syv eksperter. Som avslutning på samtalen ble ekspertene spurt om de opplevde å få tilstrekkelig informasjon i forkant, hvordan samtalen opplevdes og om den tekniske gjennomføringen og verktøyet som ble benyttet fungerte godt.

Forberedelse er viktig

Ekspertene ga tilbakemelding om at de opplevde at informasjonspakken som ble sendt i forkant var tilstrekkelig for å kunne svare på spørsmålene. En ekspert trakk frem at det kunne vært nyttig med noe bedre forklaringer av metoden og det tekniske verktøyet i informasjonspakken. I hvert av de individuelle intervjuene ble det brukt tid innledningsvis på å illustrere metoden og den tekniske løsningen.

Flertallet av ekspertene pekte på at selv om de fikk tilsendt informasjonspakken i forkant, hadde de ikke brukt nok tid på å sette seg inn materialet. Flere oppga at de mener metoden fungerer best når en ekspert har brukt tid på å gå gjennom relevante studier og forstå problemstilling i forkant, noe ikke alle hadde rukket eller prioritert. Vår oppfatning var at desto mer tid eksperten hadde brukt på forberedelse, jo raskere gikk gjennomføring av det individuelle intervjuet. Ekspertene reflekterte selv over at de til daglig har lite tid til å sette seg ned med denne typen informasjonsarbeid, men at det var viktig å

prioritere. En ekspert foreslo at rekruttering foregår gjennom seksjonsledere ved aktuelle helseforetak, som kan hjelpe med å rekruttere riktige eksperter, samt bidra til at de har tid til å prioritere deltagelse.

Positive tilbakemelding på teknisk verktøy og gjennomføring

Flertallet oppga at den tekniske løsningen var enkel å forstå og at det fungerte godt at den som intervjuet plasserte brikkene i rutenettet. I denne piloten ble tre intervjuer gjennomført fysisk og øvrige digitalt. Det var ingen forskjell i tilbakemeldingene avhengig av hvordan intervjuet ble gjennomført. En ekspert trekker imidlertid frem at verktøyet oppleves noe utilgjengelig og at det er viktig at man sørger for at alle ekspertene har full forståelse for verktøyet, for å ikke tilføre utilsiktet usikkerhet rundt estimatene.

Flere ga tilbakemelding om at de opplevde at det var nyttig å gå frem og tilbake underveis i samtalen, for å se de ulike anslagene i sammenheng. Derfor opplevdes det nyttig at man fortløpende la valgte fordelingen inn i slide-dekket man jobbet med, slik at det var lett tilgjengelig å gå tilbake.

Ekspertene kunne fritt bruke så mange brikker de ønsket, men det ble anbefalt å bruke om lag 20 brikker. Flere pekte på at det var viktig at de selv kunne velge hvor mange de skulle bruke, men at det var fint med en anbefaling.

4.2 Konsensusmøtet

I konsensusmøtet ble ekspertene samlet for å diskutere de samme spørsmålene de ble stilt i de individuelle samtalene. Alle ble presentert for de individuelle anslagene som ble gitt, og ble bedt om å enes om et anslag. Konsensusmøtet ble i denne piloten, som nevnt, gjennomført med mulighet for både fysisk og digitalt oppmøte. Tre av fem ekspert møtte fysisk. To eksperter hadde ikke anledning til å delta på konsensusmøtet.

Utfordrende med hybrid møte-gjennomføring

En viktig tilbakemelding om den praktiske gjennomføringen av konsensusmøte var at det trolig ville fungert bedre om alle deltagerne hadde deltatt digitalt, eller alle fysisk. Kroppsspråk er viktig når man som gruppe forsøker å komme til enighet, og denne dynamikken fungerte mindre bra i et hybridmøte.

Flertallet av ekspertene oppgir imidlertid at de var fornøyde med deltagelse i konsensumøtet og flere oppgir at det styrker kvaliteten på anslagene. Diskusjonen opplevde de som nyttig, oppklarende og relevant.

Det var noe delte tilbakemeldinger i om det kunne være nyttig om det deltok andre i konsensumøtet som kunne bidratt med objektive betraktninger. Dette kunne for eksempel være en statistiker eller helseøkonom, som kan løfte frem diskusjonsmomenter og sørge for en korrekt felles forståelse av hva man diskuterer. Noen oppgir at det holder med god møteledelse av en som har god kjennskap til feltet og metoden.

4.3 Bruk av metoden i praksis

Innhenting av strukturerte innspill fra eksperter krever mer arbeid enn en enkelt ekspertvurdering. Vi spurte derfor ekspertene om når de opplever det er riktig å prioritere ressurser på en mer systematisk innhenting av deres innspill. I tillegg ba dem svare på hva de opplever er viktig å tenke på i den praktiske gjennomføringen, som antall eksperter man bør ha, geografisk variasjon og eventuelle kilder til usikkerhet.

Flere opplevde at det var særlig nyttig å reflektere rundt usikkerhet i anslagene og at det er viktig at dette kommuniseres. For ekspertene opplevdes det som en nyttig øvelse å ikke kun gi et anslag, men kunne reflektere rundt usikkerhet og variasjon mellom pasienter. Det ble pekt på at denne er en utfordring når de konsulteres i dag, og at det er krevende for dem å kommunisere om de er trygge på anslagene de gir eller ikke.

Flere av ekspertene trakk frem at det er viktig at man gir gode instruksjoner og sørger for forståelse for hva man oppgir usikkerhet om, for at man skal ha tillitt til anslagene som gis.

Flere ga også tilbakemelding om at det først og fremst er viktig med et sterkt fagmiljø for å sikre representativitet i anslagene, selv om også geografisk variasjon er ønskelig. En ekspert foreslår at det holder at hvert regionssykehus eller universitetssykehus er representert. En annen trekker frem at det er nyttig å ha eksperter av begge kjønn, og også både unge og mer erfarne leger representert.

Figur 3: Oppsummering av innspill fra ekspertene

Tilbakemeldinger fra eksperter



Høyt engasjement



Konsensumøte oppleves både som en gulrot og en kilde til mindre usikkerhet



Positive tilbakemeldinger på teknisk verktøy og gjennomføring

Suksessfaktorer for gjennomføring



Erfaring er positivt

tidligere erfaring med strukturerte klinikerinnspill letter gjennomføringen



Forberedelse er viktig

tilstrekkelig planlegging og veiledning forbedrer kvaliteten og konsistensen i prosessen



Det trengs flere pilot-studier

ytterlige testing vil bidra til å forbedre rammeverke og støtte en bredere implementering

5. Drøfting

Denne piloten har gitt nyttig lærdom i den praktiske gjennomføringen av innhenting av strukturerte ekspertvurderinger. Videre oppsummerer vi disse læringspunktene. Vi trekker også frem områder vi har identifisert der det fortsatt behøves videre forskning og utprøving. Disse områdene bør være fokusområde for videre pilotering.

5.1 Antall eksperter og rekruttering

Det er fortsatt behov for bedre forståelse av hvor mange eksperter som bør inngå når denne metoden benyttes. I tilfeller med veldig sjeldne sykdommer kan det være få eksperter, særlig i et land som Norge med en liten befolkning, og dermed krevende å rekruttere nok eksperter til informasjonsinnhenting. Det kan i slike tilfeller være aktuelt å vurdere behovet for mange eksperter, og om det eventuelt er mulig å rekruttere eksperter fra andre land. Et annet alternativ vil være å benytte eksperter med noe mindre klinisk erfaring med pasientpopulasjonen.

Med intervjuer av syv eksperter fikk vi noe variasjon i anslagene, men det gjenstår å gjennomføre ytterligere analyser for å vurdere hvorvidt færre eksperter kunne gitt like verdifull informasjon.

Oakley et al (2025) slår fast at det foreløpig ikke er konsensus knyttet til det nødvendige antallet eksperter. MRC-protokollen anbefaler minst fem eksperter, uten at det gis en begrunnelse (Bojke, et al., 2021). Videre slår de fast at det viktigste er at de relevante fagområdene er dekket, og at det må være nok eksperter til at de som skal bruke informasjonen har tillitt til anslagene. Som beskrevet er tidligere ga ekspertene i denne piloten tilbakemelding om at det viktigste er at de sentrale fagområdene er dekket.

Ved bruk av metoden i praksis bør det etableres en protokoll for hvordan eksperter skal rekrutteres. I denne piloten ble metode for rekruttering ikke evaluert.

5.2 Forberedelse av eksperter

Vår vurdering er at det avgjørende at ekspertene er godt forberedt før man gjennomfører både det individuelle intervjuet og konsensusmøtet.

Ekspertene i denne piloten ga tilbakemelding om det samme, og flere trakk frem at de tror øvelsen hadde vært nyttigere dersom de hadde vært bedre forberedt.

I denne piloten ble informasjonspakken sendt ut to ganger; en gang i forbindelse med at intervjuet ble avtalt og en gang om lag to dager før intervjuet ble gjennomført. I den andre e-post ble det understreket at det er nyttig om informanten rakk å forberede seg. Dette tiltaket virker imidlertid ikke å ha stor nok effekt, og andre tiltak for å sikre gode forledelser bør vurderes.

Andre mulige metoder for forberedelse kan være å bruke mer av intervjuet helt spesifikt på å selv forberede ekspertene. Man kan også tenke seg at behovet for forberedelser faller etter hvert som ekspertene har deltatt i tilsvarende intervjuer. Oakley et al (2025) slår imidlertid fast at protokollene legger opp til at eksperter ikke skal ha erfaring med statistikk og rammeverket, og det er dermed viktig at man ikke legger opp til en gjennomføring som bygger på en forventning til tidligere kjennskap.

Videre er det vår opplevelse at det både i individuelle intervjuer og under konsensusmøtet tidvis var utfordrende for ekspertene å sørge for at de svarte for riktig pasientpopulasjon, og da særlig om anslagene de ga var tilstrekkelig tilpasset norsk klinisk praksis. Tidligere studier i pasientpopulasjonen ble i stor grad benyttet som utgangspunkt for ekspertene, både under individuelt intervju og konsensusmøte.

I denne piloten, og trolig i de fleste tilfeller det er relevant å bruke denne metoden, er det snakk om en liten pasientgruppe. Det vil si at selv om de rekrutterte er eksperter på området, vil mange ha behandlet få pasienter tidligere. Det er viktig at de balanserer funn fra studier og en generell tilpasning til forventet norsk populasjon og egne erfaringer med pasientene. De fleste ekspertene hadde behandlet færre en fem pasienter med kjent riktig mutasjon (delvis fordi man tidligere ikke teste for genforandringer i mangel på tilpasset behandling).

Vi opplevde at ekspertene var tryggere i sine anslag når de skulle anslå en kontrollarm for behandlingen, fordi dette i stor grad er basert på egne erfaringer fra klinikken, kombinert med funn fra litteraturen som slår fast prognoser for pasienter med FGFR2+ sammenlignet med pasienter uten. For intervensjonsarmen gjorde vi oss relativt like erfaringer; basert på FIGHT-202 opplevdes

ekspertene relativt trygge i sine anslag. De baserte anslagene i stor grad på forventet effekt av endrete pasientkarakteristikker i norsk populasjon sammenlignet med populasjonen i studien. Alle ekspertene forventet at pasienter i klinikken vil være noe sikrere enn pasientene i en studiesetting. Det opplevdes imidlertid som at ekspertene var mer usikre når de skulle anslå virkninger av å innføre intervensjonen i førstelinjebehandling. For langtidsvirkninger baserte ekspertene seg også på generelle betraktninger rundt behandlingseffekt med immunterapi og målrettet behandling.

En annen viktig del av forberedelsen handler om hvilken informasjon ekspertene får i forkant og bruker underveis i intervjuet. I denne piloten var FIGHT-202 og FIGHT-302 viktige studier, som vi opplevde at ekspertene i stor grad lente seg på i sine anslag. Vi opplevde imidlertid at det var nyttig å ha disse studiene lett tilgjengelig både under individuelt intervju og konsensummøte, fordi ekspertene brukte resultater fra disse studiene som utgangspunkt. Derfor anser vi det som nyttig at alle relevante studier inkluderes. Det kan gjerne lages oppsummeringer som kan inkluderes som bakgrunnsinformasjon i møtene.

Tilstrekkelig forberedelse er et viktig tiltak for å sikre forståelse for oppgaven og for at ekspertene skal kunne bruke den kunnskapen de har til å gi de mest riktige anslagene de har. Det er imidlertid utfordrende å sørge for at alle ekspertene faktisk prioriterer forberedelse, og derfor oppleves noen andre tiltak som viktig. For det første må det settes av tid i starten av møtet til å forklare rammeverket og pasientpopulasjonene. Deretter opplevde vi det som nyttig å gi eksempler på tolkning av ulike fordelinger, før man startet gjennomgangen. I tillegg opplevde vi at det kunne gi bedre forståelse om man for hvert anslag ekspertene ga oppsummerte kort hva anslaget medfører med tanke på usikkerhet. I tillegg minnet vi for hvert anslag ekspertene på hvilken populasjon det var snakk om.

5.3 Gjennomføring av konsensummøte

Konsensummøtet slik det ble gjennomført i denne piloten var i hovedsak vellykket. Det er imidlertid noen læringsmomenter vi tar med videre i arbeidet.

Ikke optimalt med hybrid gjennomføring

For det første er vår vurdering, i likhet med ekspertene, at hybride møter ikke burde gjennomføres, av flere årsaker. Hybrid gjennomføring gjorde det for det første noe vanskeligere å sørge for at alle ekspertene deltok

likeverdig i diskusjoner. Oakley et al (2025) trekker frem at det er viktig å sørge for at ikke noen ekspertene dominerer diskusjonen. Dette kan skje for eksempel hvis ekspertene har et forhold til hverandre, som at noen har mer erfaring enn andre. I et hybrid møte var det lettere for ekspertene i rommet å delta i diskusjonen. Møteleder må sørge for å være tydelig i at alle skal delta likeverdig og bidra til at alle får mulighet til å si sine meninger, noe som var utfordrende i en slik setting.

Digital deltagelse har fordeler ved at ekspertene slipper å reise for å delta. Også Oakley et al (2025) finner at digitale møter kan være vel så effektivt som fysiske møter, noe som underbygger at det kan være mer hensiktsmessig å planlegge for fulle digitale møter. Det gjør det enklere for alle å se hverandre og ha oversikt over deltagerne, sammenlignet med et hybrid møte. I tillegg blir det lettere å dele innhold på en skjerm som alle ser like godt.

Presentasjon av anslag fra individuelle intervjuer

Som tidligere nevnt presentere vi i konsensummøtet et gjennomsnitt av anslagene gitt i individuelle intervjuer. Hensikten med dette var i hovedsak å ha et utgangspunkt for diskusjon, fremfor å kun vise alle fordelingene hver for seg. Vår vurdering er imidlertid at det kan virke som at dette gjennomsnittet hadde en slags anker-effekt, der ekspertene i stor grad ble enige om dette gjennomsnittet. For flere anslag var det større variasjon i hva ekspertene hadde oppgitt i intervjuer, og det ble i liten grad diskusjoner om det kanskje heller burde vært riktig at anslaget lå lengre til en av sidene, og om man burde fanget opp alle ekspertenes anslag i et usikkerhetsspenn. Et videre spørsmål blir dermed om det er mer hensiktsmessig å kun vise alle de individuelle anslagene. Vår vurdering er imidlertid at det er pragmatisk og effektivt å starte diskusjonen med et forslag til konsensus. Et alternativ kan være å lage et forslag heller rundt ett av ytterpunktene for å få en diskusjon om det er riktig «side» å ligge på eller benytte en bredere fordeling rundt gjennomsnittet, slik at spredning i alle individers anslag er dekket.

Inkludering av andre fagpersoner

Det siste poenget er knyttet til den tidligere diskusjonen om hvorvidt ekspertene i stor nok grad svarer for riktig populasjon; om de svarer for en studiesetting eller hva som forventes i norsk klinisk praksis. Under konsensummøtet opplevdes dette som særlig utfordrende, fordi det var mer utfordrende å justere og forklare fortløpende da det var mange som deltok i diskusjonen. Ekspertene trekker frem at de er fornøyde med møteledelsen,

men at dette også er svært viktig for å lykkes med rammeverket. Et vurderingspunkt var også om det kunne vært hensiktsmessig med en annen fagperson i panelet; som en statistiker eller helseøkonom. Vår vurdering er at dette kunne vært nyttig; denne fagpersonen må både ha god forståelse for den statistiske delen, samt hvordan anslagene skal anvendes i praksis. Vår vurdering er at en slik kompetanse kunne bidratt med å øke forståelsen av anslagene, særlig betydningen av usikkerheten. Av den grunn opplever vi at det kan være mest hensiktsmessig med en helseøkonom, med erfaring innen helseøkonomisk modellering og beslutningsprosesser.

Dersom man ønsker å inkludere en annen fagperson, er det viktig at denne personen kun bidrar med objektive vurderinger i møtet, uten å påvirke eksperten den ene eller den andre veien. Vi mener derfor at det kan være interessant å utarbeide et sett med faste spørsmål og momenter denne personen kan bidra med for hvert anslag, for å sørge for at deltagelsen blir nøytral.

Bruk av Kaplan-Meier plot i konsensusmøte

Som beskrevet i kapittel 2, utarbeidet vi fortløpende Kaplan-Meier plot, som ble vist i slutten av konsensusmøtet. Dette var lagt opp som en eksplorativ tilleggsmodul etter at standard-delen av møtet var ferdigstilt. Vi opplevde at dette var en nyttig validering av ekspertenes anslag, i tillegg til at ekspertene selv fikk bedre forståelse når de så hva anslagene innebar og så de ulike anslagene i sammenheng. Dermed bidro disse analysene til å tydeligere kommunisere effekt av usikkerhet og fordelingene de har oppgitt, særlig fordi det virker som at ekspertene i større grad har erfaring med å

forholde seg til Kaplan-Meier plott enn fordelingen slik den lages i rammeverket. Det er likevel verd å merke seg at det ikke ble ønsket eller gjort noen endringer som følge av at disse kurvene ble vist. Det er ikke gitt at det metodisk er nødvendig å inkludere disse analysene.

5.4 Fokus for videre forskning

Metodene for systematisk innhenting av ekspertvurderinger er fortsatt under utvikling og det er behov for mer praktisk erfaring med gjennomføringen. Vår erfaring fra arbeidet med metoden er at det fortsatt er en rekke forhold som bør undersøkes nærmere. Dette inkluderer blant annet:

- Hvilke praktiske grep kan gjøres for at eksperter gir de beste anslagene og usikkerhetsestimatene?
- Antall punkter man innhenter informasjon om
- Empirisk validering av resultater
- Bruk av metoden på andre utfall enn PFS og OS (for eksempel dosering, varighet av behandling, prevalens)
- Konsensus vs. matematisk aggregering.
- Hvor mange eksperter bør inkluderes og hvordan bør de rekrutteres?
- I hvilke tilfeller er rammeverket særlig relevant?
- Kan rammeverket brukes på ulike sykdomsområder?
- Hva er verdien av systematisk innhenting av ekspertvurdering for beslutningsgrunnlaget?

Ytterligere testing vil bidra til å forbedre rammeverket og støtte en eventuell bred implementering i praksis.

6. Referanser

Bojke, L. et al., 2021. Developing a reference protocol for structured expert elicitation in health-care decision-making: a mixed-methods study. *Health Technology Assessment*, Juni, 25(37), pp. 1-124.

Cooke, R., 1991. *Experts in uncertainty: opinion and subjective probability in science*, USA: Oxford University Press.

Dalkey, N. & Helmer, O., 1963. An experimental application of the Delphi method to the use of experts. *Management Science*, 9(3), pp. 458-467.

Hemming, V. et al., 2018. Eliciting improved quantitative judgements using the IDEA protocol: A case study in natural resource management. *PLoS One*, 13(6).

Meld. St. 21, 2024-2025. *Helse for alle - Rettferdig prioritering i vår felles helestjeneste*.

Oakley, J. E. & A, O., u.d. *SHELF: the Sheffield Elicitation Framework (version 4)*. [Internett] Available at: <https://shelf.sites.sheffield.ac.uk/>

Oakley, J. E. et al., 2025. *NICE DSU Technical Support Document 26: Expert elicitation for long-term survival outcomes*, s.l.: s.n.

A decorative diagonal line runs from the top-left corner towards the bottom-right corner, separating the white background from a light blue area.

oslo**economics**

www.osloeconomics.no

E-post og telefon:
post@osloeconomics.no
+47 21 99 28 00

Besøksadresse:
Klingenberggata 7A
0161 Oslo

Postadresse:
Postboks 1562 Vika
0118 Oslo