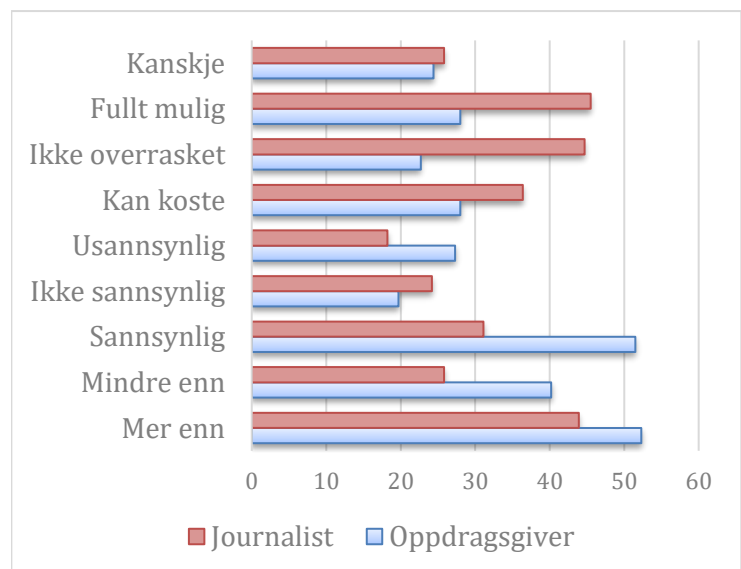


Karl Halvor Teigen, Bjørn Andersen og
Sigurd Lerkerød Alnes

Hvordan oppfattes og omtales usikre kostnadsestimer?



Arbeidsrapport

2018-3

Forord

Arbeid med denne rapporten er støttet med midler fra Concept-programmet ved NTNU til forskningsprosjektet «Kommunikasjon av kostnadsestimater med usikkerhet» som er gjennomført av forskere ved NTNU, Universitetet i Oslo og Simulasenteret.

Vi takker Gro Hege Haraldsen Nordbye, Fredrik Svartdal Færevaaag, Erik Løhre og Jan-Ole Hesselberg for deres bidrag til rekruttering av deltagere og innhenting av data, og Hilde Mentzoni Andersen for oversettelse av spørreskjema.

Trondheim, mai 2018

Karl Halvor Teigen, Bjørn Andersen og Sigurd Lerkerød Alnes

Ansvaret for informasjonen i rapportene som produseres for Concept-programmet ligger hos forfatterne. Synspunkter og konklusjoner står for forfatternes regning og er ikke nødvendigvis sammenfallende med Concept-programmets syn.

Innhold

FORORD	1
SAMMENDRAG	3
1 INNLEDNING	5
1.1 BAKGRUNN	5
1.2 BEDØMMELSER UNDER USIKKERHET	6
1.3 PILOTSTUDIER	8
1.3.1 <i>Pilot 1: Anslag av sannsynligheter</i>	8
1.3.2 <i>Pilot 2: Vurdering av estimaters riktighet</i>	9
1.4 HOVEDUNDERSØKELSEN.....	9
1.4.1 <i>Anslag</i>	10
1.4.2 <i>Vurderinger</i>	10
1.4.3 <i>Verbale sannsynligheter</i>	10
2 METODE	12
2.1 DELTAKERE OG PROSEDYRE	12
2.2 HOVEDSKJEMA (GRUPPE 1-3).....	12
2.3 FORENKLET SKJEMA TIL STUDENTER (S-SKJEMA).....	14
2.4 SPØRSMÅL TIL DOMMERGRUPPEN (D-SKJEMA).....	14
3 RESULTATER	16
3.1 RESULTATER FRA HOVEDSKJEMA (GRUPPE 1-3).....	16
3.1.1 <i>Anslag av min og maks-verdier (Q1)</i>	16
3.1.2 <i>Sannsynligheter for min og maks-verdier (Q6)</i>	16
3.1.3 <i>Annet prosjekt (Q10 og Q11)</i>	17
3.1.4 <i>Vurdering av estimatene (Q8)</i>	18
3.1.5 <i>Verbale sannsynligheter</i>	18
3.2 RESULTATER FORENKLET SKJEMA (GRUPPE 4).....	21
3.2.1 <i>Verbale sannsynligheter</i>	21
3.2.2 <i>Vurdering av estimatene</i>	21
3.3 RESULTATER DOMMERE (GRUPPE 5)	22
3.3.1 <i>Verbale sannsynligheter (DQ1)</i>	22
3.3.2 <i>Vurdering av estimatene (DQ2)</i>	22
4 DISKUSJON	24
4.1 SNEVRE INTERVALLER OG FORVENTNINGER OM OVERSKRIDELSER	24
4.2 VURDERING AV ESTIMATER: EN GRENSEEFFEKT	24
4.3 VERBALE SANNSYNLIGHETER: EN EKSTREMITETSEFFEKT	25
4.4 BEGRENSNINGER	26
5 KONKLUSJON	28
REFERANSER	30

Sammendrag

Som underlag for beslutninger om konseptvalget i store offentlige investeringsprosjekter utarbeider eksterne rådgivere et kvalitetssikringsdokument (KS1) som gjennomgår de mest aktuelle alternativer, sammen med et kostnadsestimat for hvert av disse, og senere (KS2) en kvalitetssikring av interessentanalyse, prosjektstrategi, kontraktsmodell og også usikkerhetsanalyse. Disse bygger på kostnadsestimater med usikkerhetsvurderinger utarbeidet av etatene. På grunn av den usikkerhet som alltid er til stede tar estimatene form av en sannsynlighetsfordeling hvor mest sannsynlig (i symmetriske fordelinger også kalt forventet) kostnad blir oppgitt, og i tillegg også minimums og maksimums-verdier, definert som P15 og P85 i en kumulativ sannsynlighetsfordeling. Disse estimatene vil ikke bare være tilgjengelig for oppdragsgiver, men også for andre interesserte lesere som ønsker å danne seg et bilde av «prisappen» på et prosjekt og usikkerheten forbundet med de ulike alternativene.

I denne undersøkelsen tar vi for oss hvordan lesere som ikke er direkte involvert i estimeringsprosessen tolker, omtaler og vurderer slike estimater, med utgangspunkt i et kostnadsestimat fra KS1-rapporten for et faktisk veiprojekt. Vi vet fra annen forskning at folks tenkning om sannsynligheter og usikkerhetsintervaller ikke alltid faller sammen med disse begrepens matematiske implikasjoner. I tillegg formidles slike estimater verbalt, med ord og uttrykk (usikkert, stor sjanse, mulig) som ikke lett kan tilordnes til tall. Når brukes de og hva forstår vi med dem? Dette er viktig fordi det i sin tur påvirker tolkningen av resultatene ex post (når er et avvik å betrakte som en skandale, et svært feilaktig estimat, osv.).

Rapporten presenterer svarene på et spørreskjema, som i litt forskjellige varianter ble gitt til fem utvalg med ulik faglig bakgrunn: Fagfolk i et stort entreprenørfirma (N = 48), studenter på et kurs om prosjektstyring ved NTNU (N = 51), bachelorstudenter på kurs om beslutnings- og bedømmingspsykologi ved UiO (N = 36), førsteårsstudenter i psykologi (N = 134) og dommere på kurs arrangert av domstolsadministrasjonen (N = 256).

På tross av ulikheter mellom gruppene i faglig innretning og grad av ekspertise er det en del fellestrekk ved svarene, som også er i tråd med resultater innhentet på andre fagfelt.

- Usikkerhetsintervallene antas å være snevrere enn de som fremkommer i KS-rapporten.
- Sannsynlighetsanslag for ulike utfall blir lett i overkant store. På tross av snevre intervaller vurderes spesielt sjansene for overskridelser som store.
- Estimater blir bedømt som korrekte når kostnadene svarer til maksverdien og ukorrekte når de ligger over, selv om forskjellen kan være relativt beskjeden (og sannsynlighetsfordelingen innebærer at overskridelser er mulige). Denne grenseeffekten er likevel mindre enn i studier hvor deltagerne bare tar stilling til ett utfall og ikke har tilgang til en graf over sannsynlighetsfordelingen.
- En rekke ord og uttrykk, ikke minst «muntlige» formuleringer av typen fullt mulig, ikke overraskende, og kan koste, brukes spesielt til å karakterisere høye kostnader (ekstremitetseffekten), og kan bidra til å «sannsynliggjøre» slike relativt sjeldne utfall.

Det knytter seg en god del usikkerhet til disse resultatene siden vi bare bruker ett scenario, utvalgene er relativt små og ikke nødvendigvis representative, og vi mangler data fra viktige grupper, som involverte beslutningstakere (politikere) og profesjonelle formidlere (journalister). Konklusjonene må derfor oppfattes som preliminare.

Vår studie indikerer at selv velutdannede lesergrupper i blant kan trenge ytterligere «voksenopplæring» for å forstå probabilistiske kostnadsestimater. Leverandørene av slike estimater kunne på sin side ha nytte av å kjenne noen hovedspor i lekfolks tankeverden, og derved øke sjansen til å rydde misforståelser av veien der de lettest oppstår. Det er for eksempel ikke sikkert at man bør komme leseren i møte ved å snakke om minimum, maksimum og «forventede» eller «mest sannsynlige» estimater, fordi slike betegnelser lett blir oppfattet mer konkret enn det er dekning for. En visuell fremstilling av fordelingen, som vi har brukt i denne studien, kan også være et skritt på vei til økt forståelighet, men løser ikke alle vansker.

1 Innledning

1.1 Bakgrunn

Et sentralt moment i prosjekters tidligfase og dets ferd gjennom beslutningsprosessens mange trinn frem mot finansieringsvedtak er kostnadsestimater og deres usikkerhet. For større offentlige investeringsprosjekter (med forventet kostnadsramme høyere enn 750 mill. kroner) anvendes en kvalitetssikringsordning (KS-ordningen). Finansdepartementet har inngått rammeavtale med seks konsortier av eksterne rådgivere om gjennomføring av kvalitetssikringene. KS-ordningen omfatter to kontrollpunkter i investeringsprosjektets planleggingsprosess:

- KS1 - Kvalitetssikring av konseptvalg før beslutning i regjeringen om å starte forprosjekt.
- KS2 - Kvalitetssikring av styringsunderlag samt kostnadsoverslag før eventuell investeringsbeslutning i Stortinget.

Kostnadsestimater opptrer helt fra tidlige diskusjoner om mulige tiltak innledes, er sentrale i KUVer (konseptvalgutredninger), vurderes av eksterne kvalitetssikrer i KS1 og gjennomgås i detalj i forprosjektfasen (sentralt styringsdokument) og kvalitetssikring av dette i KS2. De faglige metoder og begreper som benyttes for å få frem kostnadsestimater og fremstille den usikkerheten disse er beheftet med, er godt dokumentert i prosessdokumentasjon og veiledere utarbeidet av ulike etater og departementer, se for eksempel veileder nr. 6 om kostnadsestimering (Finansdepartementet, 2008). Det må også antas at fagpersoner involvert i både prosessen med å utarbeide estimater, kvalitetssikre disse og tolke dem ved ulike milepæler har nødvendig forståelse for disse begrepene. Imidlertid vet vi at tallene leses, tolkes og videreformidles av mange tiltenkte eller ikke tiltenkte målgrupper som kanskje ikke behersker begrepsapparatet like godt, for eksempel politikere (på ulike nivåer: regionalt, departement, Storting), saksbehandlere, journalister som formidler nyheter om prosjekter til befolkningen, “mannen i gata”, osv.

I disse dokumentene formuleres usikkerheten i kostnader knyttet til det enkelte alternativ som en kumulativ sannsynlighetsfordeling hvor P15, P50 og P85 er spesielt markert. Mens P50 (medianverdien i denne fordelingen) gjerne omtales som forventet eller mest sannsynlig kostnad (ved en symmetrisk sannsynlighetsfordeling er forventet og mest sannsynlig lik, dette vil ikke være tilfelle ved en asymmetrisk fordeling, men gjennom rapporten benytter vi disse begrepene om hverandre) blir P15 omtalt som minimum og P85 som maksimumsverdi. Disse verdiene har konsekvenser for den videre behandling av kostnadene ved at P85 ofte blir satt som kostnadsrammen for det valgte alternativ og P50 blir satt som styringsrammen.

Dette kan by på ulike problemer. Erfaringsmessig kan mange lesere av slike dokumenter tolke forventningsverdien i et kostnadsestimat som et absolutt tall eller øvre grense for kostnader, og dermed bli overrasket om sluttkostnaden overstiger dette. Det kan også være forvirrende at den etaten som står for gjennomføringen av et prosjekt normalt tildeles P50-estimatet som sin styringsramme, mens det bevilgende departementet setter av P85-estimatet som den øvre kostnadsramme for prosjektet. Differansen er en form for reserve som statistisk vil tas bruk av i

halvparten av tilfellene (mens den andre halvparten vil bruke mindre enn P50). Både media, politikere og befolkningen generelt kan oppfatte at et prosjekt som går over P50-nivået har en overskridelse, men dette er altså ikke nødvendigvis tilfelle. Situasjonen ble godt illustrert i et innlegg i Dagens Næringsliv 10. november 2016 der forskere fra SINTEF kommenterer på ulike medieoppslag rundt kostnadene for de nye kampflyene, der det falt beskyldninger om både talltriksing og endret pris på flyene (Johansen & Landmark, 2016). Her har bruken av begrepene P50 og P85 skapt forvirring når anskaffelsesprosessen har gått sin gang og kostnadene har økt. De var da langt over P50, men fortsatt innenfor P85-nivået, men dette ble av mange oppfattet som en overskridelse (dette skyldtes nok dels også at Forsvaret selv hadde brukt P50-begrepet i sitt saksfremlegg).

Studier av prosjekter som har vært gjenstand for slik ekstern kvalitetssikring siden KS-ordningen ble etablert i år 2000 viser flere tegn på positiv utvikling. Dels er det liten forskjell mellom etatenes egne kostnadsestimat og estimat fra KS2 (Welde, 2014). I gjennomsnitt var etatenes estimat 1,6 prosent lavere enn kvalitetssikrernes, men estimat i KS2 er noe nærmere sluttkostnaden enn etatenes egne estimat. Forskjellen kan også synes å ha blitt mindre over tid, den var i snitt 6 prosent ved en tilsvarende gjennomgang av de 56 første prosjektene i 2006.

Det foreliggende prosjektet har til hensikt å studere denne praksis fra et brukerperspektiv. Vi spør hvordan modellen for å kommunisere en sannsynlighetsfordeling oppfattes av lesere som ikke er direkte involvert i estimeringsprosessen, og hvorvidt det kan oppstå avvik mellom hva en KS rapport er ment å si og hvordan den kan leses av mottakere med ulik faglig bakgrunn og grad av ekspertise. Vi har valgt å spørre grupper som må antas å ha en viss forståelse av tallmessige estimater (studenter på ulike nivåer), praktisk erfaring med prosjekter (ansatte i et entreprenørfirma) og beslutninger under usikkerhet (dommere). Vår opprinnelige intensjon om også å inkludere saksbehandlere og beslutningstakere (politikere) for store offentlige investeringsprosjekter lot seg av praktiske grunner ikke gjennomføres innenfor tidsrammen for dette prosjektet.

Mer spesifikt har vi undersøkt tre forhold: Hvordan folk oppfatter spennet i usikkerhetsfordelingen (avstanden mellom minimums og maksimumsverdi, og sannsynligheten for disse verdiene), hvordan de vurderer utfall som faller innenfor eller utenfor dette spennet, og hvordan slike utfall kan bli karakterisert verbalt.

Det finnes en omfattende forskningslitteratur med relevans for alle disse tre spørsmålene innenfor bedømmings- og beslutningspsykologien, samt innenfor en del utvalgte spesialområder (f.eks. klimaforskning og etterretningsanalyse) hvor tolkning og rapportering av usikkerheten står sentralt. Vi vil i det følgende kort nevne noen bidrag fra denne litteraturen som kan være aktuelle også ved formidlingen av usikre kostnadsestimater.

1.2 Bedømmelser under usikkerhet

Atskillig forskning viser at lekfolk og i blant også eksperter tenker annerledes om sannsynligheter og estimater av usikre verdier enn det er statistisk grunnlag for. For eksempel blir summen av anslåtte sannsynligheter for et uttømmende sett av alternative utfall ofte langt over 100%, også omtalt som «sub-additivity» (Tversky & Koehler, 1994) og «additivity neglect» (Riege & Teigen,

2013). Leger som skulle angi sjansene for fem ulike utfall for en pasient med hjerteinfarkt kom opp i summer på gjennomsnittlig 169% (Redelmeier, Koehler, Liberman & Tversky, 1995). Studenter som skulle vurdere vindersjansene for samtlige låter i Melodi Grand Prix i 1986 oppnådde summer på gjennomsnittlig 555%, nok til å kåre fem og en halv vinner (Teigen, 1988). Konfrontert med slike misforståelser svarer de gjerne at de har vurdert sjansene for hvert enkelt alternativ «hver for seg», og ikke innsett at sannsynligheten for ett utfall har konsekvenser for hvor store sjansene er for de øvrige. Vi kan altså ikke uten videre ta det for gitt at en leser som får vite at sannsynligheten for et intervallestimat er 70% skjønner at det bare er igjen en 30% sjanse for utfall under minimum og over maksimum tilsammen. Mange skjelner også dårlig mellom sannsynligheten for et punkttestimat og for intervallet rundt det.

Anslåtte konfidensintervaller for hva som kan sies med f. eks. 90% eller 95% sikkerhet blir typisk alt for snevre, gjerne omtalt som overkonfidens (Moore & Healey, 2008) eller overpresisjon (Moore, Tenney & Haran, 2016). Lekfolk kan ha vanskeligheter med å innse at et 90% konfidensintervall nødvendigvis må bli videre enn et 60% intervall (Langnickel & Zeisberger, 2016; Teigen & Jørgensen, 2005), og vurderer ofte snevre intervaller som mer troverdige og sikrere enn vide (Løhre & Teigen, 2017).

Uansett grad av sannsynlighet blir intervallene lett oppfattet som kategorier, hvor utfall som faller innenfor blir vurdert som korrekt anslått og utfall utenfor blir vurdert som feilestimert. En studie av klimarelaterte prognoser viste at en ekspert som predikerte flom på 2-9 km med 90% sikkerhet ble oppfattet som helt korrekt ved faktisk flom på 2, 3, 8 og 9 km, mens en annen som predikerte flom innenfor et snevrere intervall 4-7 km, men da med 50% sikkerhet, ble vurdert til å ha tatt feil i sine prognoser (Teigen, Løhre & Hohle, 2018, Study 2). Det dreier seg her om en effekt av vilkårlig satte grenser for størrelsen av intervallet (boundary effect). I disse studiene var bare spennvidden for usikkerhetsintervallet angitt og ikke den fullstendige sannsynlighetsfordelingen, som jo viser at også utfall utenfor intervallet kan ventes å inntreffe fra tid til annen, riktignok med lav sannsynlighet.

Problemer med å beregne, formulere og forstå numeriske sannsynligheter gjør at de både i dagligtalen og av eksperter ofte erstattes eller suppleres av mer omtrentlige verbale uttrykk (stor sjanse, nesten sikkert, lite sannsynlig). Men forskningen på slike ord og uttrykk viser at de brukes og oppfattes svært forskjellig avhengig av situasjon og tema. De er «vage» (Budescu & Wallsten, 1995).

Det mangler ikke forsøk på å standardisere slike uttrykk innenfor bestemte fagfelt, som i klimaforskningen (Mastrandrea et al., 2010), risiko ved bruk av legemidler (EEC, 1998) og i etterretningstjenesten (Barnes, 2015; Dhami, Mandel, Mellers & Tetlock, 2015). Slike forsøk faller ikke alltid heldig ut hvis definisjonene strir mot naturlig språkbruk. For eksempel brukes «vanlig» i pakningsvedlegg for legemidler til å betegne bivirkninger som har minst 10% sjanse for å opptre, mens lekfolk gjerne tror at vanlig vil si over 50% sjanse (Berry, Raynor, Knapp, & Bersellini, 2003).

Forsøk på å tallfeste hva ulike verbaluttrykk betyr kan komme i konflikt med hvordan disse ordene brukes i praksis. På direkte spørsmål vil mange «oversette» usannsynlig med en 10-30% sjanse (etter IPCCs standard skal «unlikely» kunne brukes om alle sannsynligheter under 33%; Mastrandrea et al., 2010). Imidlertid viser det seg at det folk omtaler som usannsynlig er verdier

som i praksis ikke vil forekomme, spesielt utfall som er ekstremt høye (Teigen, Juanchich & Riege, 2013). Fenomenet har vært omtalt som en ekstremitetseffekt (Jenkins, Harris & Lark, 2018; Teigen, Filkuková & Hohle, 2018).

Språklige formuleringer er videre retningsbestemte (directional): de dirigerer oppmerksomheten enten i retning av hva som kan skje (positive, bekræftende) eller alternativt ikke skje (negative, benektende) (Honda & Yamagishi, 2017; Teigen & Brun, 1999). For eksempel kan en sannsynlighet på 40% for et utfall T uttrykkes både positivt som en viss mulighet for T, og negativt som ganske usikkert om T vil inntreffe. Ordvalget vil dermed kunne antyde anbefalinger og stillingtagen.

I tillegg vil ordene ha andre meningskomponenter som ikke like lett kan fanges opp av tall. De kan si noe om hvorvidt betingelsene for et utfall er til stede (eks.: mulig); videre om kilden til usikkerhet kan lokaliseres i den ytre verden (eks.: en sjanse) eller i manglende informasjon (eks.: uvisst), og dermed gjenspeile distinksjonen mellom såkalt aleatorisk og epistemisk sannsynlighet (Hacking, 2001; Ülkümen, Fox & Malle, 2016). I rent numeriske sannsynlighetsfordelinger for estimerte kostnader kan slike distinksjoner ikke spores, idet det gis inntrykk av at all usikkerhet, uansett kilde, er inkludert. En slik «one-size-fits-all» tilnærming er praktisk, men ikke uproblematisk (Parker & Risbey, 2018), ved å gi inntrykk av at all usikkerhet kan beregnes.

1.3 Pilotstudier

To pilotstudier ble gjennomført: Den første for å sjekke hvordan ikke-spesialister oppfatter sannsynlighetene for markerte verdier i en sannsynlighetsfordeling, den andre for å få et innblikk i hva de vil vurdere som et «riktig» eller et «feilaktig» estimat. Deltagerne i begge studiene var studenter som fulgte forelesninger i beslutning og bedømmingspsykologi ved UiT (N = 41) og UiO (N = 38). De fikk skjemaer med informasjon hentet fra en KS1-rapport om estimerte kostnader ved et veiprosjekt, illustrert med en symmetrisk klokkeformet sannsynlighetsfordeling (som avbildet på s. 11).

1.3.1 Pilot 1: Anslag av sannsynligheter

Studentene i denne gruppen fikk ett av to ulike skjemaer, A og B. Begge skjemaer informerte om at de avmerkede verdier i fordelingen svarte til 750 mill., 1100 mill. og 1450 mill. kroner. Midtverdien ble beskrevet enten som forventet verdi (Skjema A) eller som mest sannsynlig (Skjema B) siden begge disse uttrykkene brukes i KS-rapportene. Den lave og den høye verdien ble betegnet minimum og maksimum på begge skjemaer. Plasseringen av disse verdiene gikk frem av grafen, men de kumulative p-verdiene (P15, P50 og P85) ble ikke nevnt.

Spørsmålet var sannsynlighetene for slike kostnader. I Skjema A gjaldt det sannsynlighet for minimum, maksimum og forventet kostnad. I Skjema B gjaldt det sannsynlighet for kostnader under minimum, over maksimum og mest sannsynlig kostnad. De to formuleringene var valgt for å se om deltakerne skiller mellom sannsynlighet for enkeltverdier og for intervaller.

Resultater. Sannsynligheten for estimert kostnad ble anslått ganske likt på de to skjemaene, Mforventet = 55.6% og Mmest sannsynlig = 61.2%. Dette er svært høye verdier hvis man tolker estimatene til å gjelde 1100 mill. eksakt (som vel egentlig er et ganske usannsynlig utfall), og selv

om man oppfatter estimatet omtrentlig, (f.eks. som 1100 pluss minus 50 mill.) blir det neppe mer enn 15% sjanse for det.

Sannsynligheten for et minimumsestimert ble på skjema A anslått til 25.0% og under minimum (skjema B) til 13.6%. Maksimumsestimertene på begge skjemaene ble derimot anslått omtrent like høye $M_{maksimum} = 26.6\%$ og $M_{over maksimum} = 25.5\%$. Dermed forblir det uklart om de har tenkt på sannsynligheter for enkeltverdier eller for intervaller som forskjellige spørsmål eller spørsmål om det samme.

1.3.2 Pilot 2: Vurdering av estimaters riktighet

Studentene i denne gruppen fikk alle samme graf hvor minimum, forventet verdi og maksimum var markert. Denne gang var markeringene forklart som P15, P50 og P85 i en kumulativ fordeling, men bare estimatet for P50 var oppgitt (1100 mill. kroner). Spørsmålet gjaldt hva de trodde ekspertteamet hadde kommet frem til som maksimums- og minimumsestimert. Etter å ha skrevet sine anslag skulle deltagerne tenke seg at prosjektet ble gjennomført for en kostnad av 1450 mill. kroner (et beløp som vi på forhånd antok mange, men ikke alle ville oppfatte som en kostnadsoverskridelse). Gitt denne «faktiske» kostnaden ba vi deltagerne vurdere de opprinnelige estimater på en skala fra 1: helt korrekte til 9: helt gale.

Resultater. Gruppen gjettet på realistiske minimums- og maksimumsverdier (anslått til rundt 800 og 1450 mill), Deltagere som hadde oppgitt en lavere maksimumsverdi (under 1450 mill) oppfattet ekspertgruppens opprinnelige estimater som mer gale enn riktige, $M = 5.12$, mens de med maksimumsverdi over 1450 mill vurderte estimatene mer riktige enn gale, $M = 3.06$, forskjellen er signifikant $t(1, 31) = 3.66$, $p = .001$, Cohen's $d = 1.28$ (merk at høye skårer her betyr at estimatene vurderes som feilaktige). Dermed får vi en høy korrelasjon mellom anslått maksverdi og vurdert riktighet, $r = .67$ ($p < .01$).

Begge pilotstudiene var basert på små grupper uten fagkunnskap om estimering. Pilot 1 peker likevel på at vurdering og tolkning av sannsynligheter er problematisk for svært mange, i dette tilfelle studenter med bestått innføringskurs i statistikk. Pilot 2 indikerer også maksverdiens betydning for om et estimat vil vurderes riktig eller galt. Disse to temaene ble videreført i hovedundersøkelsen som beskrevet under.

1.4 Hovedundersøkelsen

I denne studien undersøkte vi hvordan ulike grupper fagfolk og legfolk oppfatter en sannsynlighetsfordeling for kostnader hvor P15, P50 og P85 er markert. Eksemplet er hentet fra en faktisk KS1-rapport om antatte kostnader ved bygging av en delstrekning av E18-E6 i Follo (Dovre Group & Transportøkonomisk institutt, 2009), som også lå til grunn for de to pilotstudiene. Dette veiprosjektet fremsto som rimelig representativt for slike utredninger og ble antatt å være ganske nøytralt (lite kontroversielt) for utenforstående lesere. KS1-rapportens Alternativ 1a, som er vårt gjennomgangseksempel i denne undersøkelsen, hadde en «mest sannsynlig» (forventet) kostnad på 1100 mill., mens P15 var satt til 750 mill. og P85 til 1450 mill. kroner. Reell sluttkostnad foreligger ikke da prosjektet siden har blitt endret, men analysen av porteføljen av prosjekter under KS-ordningen viser at et flertall faller mellom P15 og P85 og

nærmere 80 prosent av prosjektene holder seg innenfor kostnadsrammen vedtatt av Stortinget som vanligvis settes nært P85-estimatet (Welde, 2017).

Vi oppgir her forventet kostnad (P50 = 1100 mill.) og har spurt utvalgte grupper om (1) å anslå hva de tror vil være P15 og P85 for dette prosjektet og (2) hvordan de vil bedømme riktigheten av estimatene hvis faktiske kostnader faller innenfor eller utenfor disse verdiene. I tillegg spør vi (3) hvordan de vil karakterisere ulike utfall med vanlige ord som usannsynlig, fullt mulig og ikke overraskende. Bakgrunnen for disse tre temaene blir redegjort for nedenfor.

1.4.1 Anslag

Hvordan vil grupper med ulik ekspertise anslå min. og maks. estimater (P15 og P85) i en sannsynlighetsfordeling hvor slike verdier kan beregnes? Høy grad av overpresisjon (for snevre intervaller) har vært påvist i en rekke situasjoner hvor folk er bedt om å anslå 90% konfidensintervaller (svarende til P5 og P95), men i liten grad for 60% og 50% intervaller (Langnickel & Zeisberger, 2016; Moore, Tenney & Haran, 2015; Teigen & Jørgensen, 2005). Disse studiene har stort sett benyttet frekvensfordelinger (hit rates) som underlag for sannsynlighetsfordelingene. Vi er ikke kjent med forskning hvor deltakerne er bedt om å anslå andre eksperters sannsynlighetsfordelinger slik som i KS-rapportene. Spørsmålet er da om også slike usikkerhetsintervaller blir undervurdert, og om deltagerne har forståelse for hva sannsynlighetsverdier som P15, P50 og P85 innebærer. Det er f. eks. forholdsvis lett å forklare og verifisere prediksjonsintervaller for gjentatte hendelser (eks.: vintertemperaturer) (Joslyn, Nemeč & Savelli, 2013), men langt mer utfordrende for engangshendelser som kostnadsestimater for et spesifikt prosjekt. Alternativt kan det tenkes at ekspertenes sannsynlighetsfordeling er lettere å anslå fordi de også bærer preg av overpresisjon.

1.4.2 Vurderinger

Hvordan vil folk flest vurdere riktigheten av kostnadsestimater, avhengig av om faktiske kostnader faller innenfor eller utenfor markerte verdier? En studie av klima-relaterte prognoser tyder på at lekfolk oppfatter intervallprediksjoner som riktige hvis faktisk utfall ikke overstiger maksverdier, og gale om faktisk utfall faller utenfor intervallet (Teigen, Løhre & Hohle, 2018), selv med et probabilistisk konfidensintervall. De behandler altså intervallet kategorisk selv om fordelingen må antas å være kontinuerlig (f. eks. tilnærmet normal) og valget av grenseverdier er vilkårlig. Misforståelser m.h.t. slike fordelingers form er utbredt (Dieckman, Peters & Gregory, 2015; Kalinovski, Lai & Cumming, 2018). I den foreliggende studien har vi valgt å gi en grafisk (normalfordelt) illustrasjon av sannsynlighetsfordelingen med P15 og P85 tegnet inn for å unngå mistolkninger. Vi lot videre de samme deltagerne bedømme flere ulike utfall slik at de kunne sammenligne estimatene. Videre benyttet vi oss i denne undersøkelsen av deltagerens egne anslag av P15 og P85 fremfor oppgitte verdier. Dette innebærer trolig at de antatte maks og min-verdiene i mindre grad vil oppfattes som «offisielle» absolutte grenser, og vil kunne bidra til at forskjellen på kostnader innenfor og utenfor grensen blir mindre skarp.

1.4.3 Verbale sannsynligheter

Hvilke ord og uttrykk faller det naturlig å bruke i omtalen av usikre estimater? Ordvalget kan på den ene siden reflektere hvor sannsynlig man oppfatter et bestemt estimat («kostnader mellom x og y er sannsynlige»), men også estimatets kvalitet («usikkert»), ved siden av en indikasjon på kostnadens størrelse. For eksempel ser det ut til at utsagn om hva som er «mulig» eller som «kan»

inntreffe gjerne brukes til å karakterisere høye (i motsetning til middels eller lave) verdier (Teigen, Filkuková & Hohle, 2018; Teigen, Juanchich & Filkuková, 2013), mens usannsynlige utfall brukes til å karakterisere enda mer ekstreme utfall, gjerne slike som (hittil) ikke er observert (Teigen, Juanchich & Riege, 2013). Denne ordbruken kan skape kommunikasjonsproblemer hvis ordene bare oppfattes som uttrykk for grader av sannsynlighet. I denne studien ber vi deltagerne fylle ut tallverdier knyttet til uttrykk som «fullt mulig» «ikke overrasket» og «kan koste» for å undersøke om disse vil koples til ekstremverdier også i en kontekst av kostnadsanslag.

2 Metode

2.1 Deltakere og prosedyre

Et digitalt spørreskjema (Qualtrics) ble distribuert til tre grupper med varierende grad av kjennskap til prosjektplanlegging og usikkerhetsvurderinger.

- *Gruppe 1*, deltagere var 48 ansatte i et stort entreprenørfirma, (43 menn og 5 kvinner, median alder 45 år) med antatt erfaring i planlegging og gjennomføring av bl.a. veiprosjekter. Vi omtaler i det følgende denne gruppen som «fagfolk». Skjemaet ble åpnet av i alt 198 ansatte, men bare 25% besvarte.
- *Gruppe 2* besto av 51 studenter på et kurs i prosjektstyring ved NTNU (42 menn og 9 kvinner, median alder 24 år), mange med ingeniørutdanning, heretter kalt «prosjektstudenter». Disse mottok skjema oversatt til engelsk. Skjemaet ble åpnet av 92 deltagere, med en svarprosent på 52%.
- *Gruppe 3* omfattet 36 bachelorstudenter (9 menn og 27 kvinner, median alder 21 år) som deltok på kurs i bedømming og beslutningspsykologi ved Universitetet i Oslo («BB-studenter»). Alle i denne gruppen besvarte skjemaet.

Utvalgte spørsmål fra hovedskjemaet ble i tillegg stilt til de følgende to grupper, som av tidshensyn og andre praktiske grunner ikke kunne besvare et fullstendig skjema:

- *Gruppe 4*. Begynnerstudenter. En forenklet papirutgave av spørreskjemaet ble fylt ut av to grupper førsteårs-studenter i psykologi i Lillehammer (N = 64) og i Oslo (N = 70), 68% kvinner, median alder 21 år.
- *Gruppe 5*. To tilpassede spørsmål ble inkludert i et elektronisk spørreskjema distribuert til dommere før deltagelse på kurs i bedømmingspsykologi arrangert av domstolsadministrasjonen (Hesselberg, 2018). Av 395 deltagere som åpnet skjema besvarte 356 (85%) våre spørsmål, 46% kvinner, median alder 54 år. Som yrke oppga 182 tingrettsdommer, 89 lagdommer, 5 høyesterettsdommer, 78 annet (2 uoppgitt).

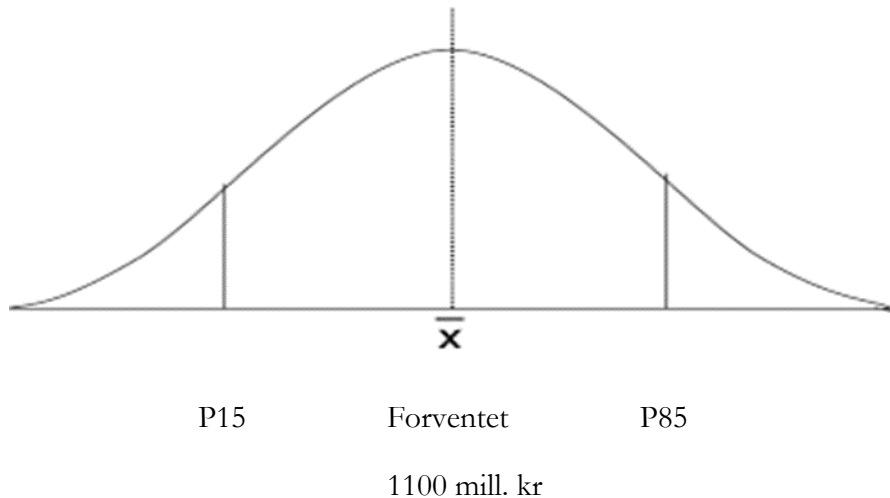
Studentene i Gruppe 2, 3 og 4 besvarte skjemaet i tilknytning til forelesning. mens deltagere i Gruppe 1 og 5 mottok en lenke til utfylling etter eget ønske. Skjemaet inneholdt ikke demografisk informasjon ut over kjønn, alder og yrke/studium. All deltagelse foregikk på frivillig basis og var ikke honorert.

2.2 Hovedskjema (Gruppe 1-3)

Spørreskjemaet til alle gruppene åpnet med en kort orientering om Concept-programmet og kvalitetssikring av store offentlige prosjekter før byggestart. De ble fortalt at kvalitetssikringen omfattet en usikkerhetsanalyse av hvert enkelt alternativ hvor en ekspertgruppe beregner forventet eller mest sannsynlig kostnad sammen med et minimums og et maksimumsestimert.

Disse verdiene ble kort forklart og illustrert med en symmetrisk klokkeformet kurve hvor P15, P50 og P85 var tydelig markert (se nedenfor).

Forventet kostnad for det anbefalte alternativ var oppgitt til 1100 mill. I den elektroniske versjon av skjemaet var verdiene for P15 og P85 var ikke oppgitt.



Grafen ble fulgt av 11 spørsmål oppsummert i Tabell 1.

Tabell 1. Spørsmål på hovedskjema (til Gruppe 1-3)

Nr	Spørsmålstekst
Q1	Hva tror du teamet anslo som minimum- og maksimumsverdier?
Q2	(Bare til Gruppe 1) Tenk deg at du skal gi en ikke-teknisk beskrivelse av hva denne veien vil koste og usikkerheten forbundet med dette. Formuler to-tre setninger uten å oppgi numeriske sannsynligheter (bruk isteden ord som «stor sjanse», «liten sjanse» «mulig», «usikkert» osv.).
Q3	Tenk deg en uformell samtale mellom kolleger som har lest prosjektrapporten. Fyll ut disse setningene med et kostnadsbeløp slik at de gir mening og virker naturlige i denne sammenhengen. I alt 11 utsagn ble presentert i randomisert rekkefølge. Prosjektet vil koste mer enn / mindre enn / sannsynlig / ikke sannsynlig / fullt mulig / kan koste / vil kanskje koste / jeg blir ikke overrasket om det koster mill. kr.
Q4	Hvilke tre av disse utsagn vil du bruke i en samtale med oppdragsgiver?
Q5	Sett at samtalen blir overhørt av journalist fra Finansavisen. Hvilke tre utsagn tror du han vil notere seg for bruk i avisen?
Q6	Sett at teamet som kom frem til de opprinnelige anslagene var spurt om hvor sannsynlig de selv vurderte dem (som tall mellom 0 og 100%), hva tror du de ville svare? <i>Gruppe 1:</i> Forventet kostnad (± 50 mill.), minimumsverdi (± 50 mill.), maksimumsverdi (± 50 mill.).

	<i>Gruppe 2-3:</i> Under min (P15); mellom min og maks; over maks (P85). For halvparten av deltakerne i Gruppe 2 var grafen over sannsynlighetsfordelingen fra innledningen gjentatt ved dette spørsmålet.
Q7	Sett at prosjektet blir godkjent av oppdragsgiver i henhold til planene. Ved prosjektstart utarbeides en ny rapport (KS2) av en tilsvarende ekspertgruppe som den som kom frem til en kostnad på 1100 millioner. Det foreligger nå mer oppdatert og detaljert informasjon om hva kostnadene vil bli. Hva tror du de nye anslagene vil være? (Ny forventet kostnad/ ny minimumskostnad / ny maksimumskostnad).
Q8	Sett at det viser seg at kostnadene kom på maksimum / maks – 100 mill. / maks + 100 mill. Hvordan vil du vurdere estimatene? (skala fra 1: helt feil til 7: helt riktig). I de digitale skjemaene var «maks» her automatisk fylt ut med det anslag den enkelte deltager selv hadde oppgitt i Q1.
Q9	Hvis sluttkostnad oversteg mill. kr ville det vurderes som et skandaleprosjekt
Q10	I et annet veiprosjekt vedtatt for oppstart er den forventede kostnaden 2500 mill. Hva vil du anbefale at Vegvesenet får som sitt budsjett for prosjektet?
Q11	Hvor sannsynlig (i prosent) anser du det for å være at sluttkostnaden for veien blir under 2500 mill.?

2.3 Forenklet skjema til studenter (S-skjema)

Til Gruppe 4 (begynnerstudenter) var det utarbeidet et forenklet spørreskjema til utfylling på papir. Her var min og maks oppgitt til 750 mill. og 1450 mill. kroner (faktiske verdier fra KS1) og inntegnet i den normalfordelte grafen. For halvparten av deltagerne ble disse verdiene kalt lavt og høyt anslag i stedet for min og maks, for å motvirke et inntrykk av et intervall med skarpe grenser.

Skjemaene inneholdt følgende spørsmål.

- SQ1 Grafen var fulgt av et spørsmål om hvilke kostnader deltagerne vil kalle sannsynlige og hvilke de vil kalle usannsynlige.
- SQ2 På neste side var figuren gjentatt og deltagerne fylte ut fire utsagn med passende kostnadsbeløp (svarende til Q3 i Tabell 1); Prosjektet kan koste / vil koste mer enn / Jeg blir ikke overrasket om det koster / det er fullt mulig at det vil koste Mill. kr.
- SQ3 Sett at det viser seg at kostnadene kom på 700 mill. / 800 mill. / 1400 mill. / 1500 mill. Hvordan vil du vurdere estimatene? (skala fra 1: helt feil til 7: helt riktig). Spørsmålet svarer til Q8 i hovedskjemaet, men beløpene var valgt for å illustrere utfall like under og like over oppgitt minimum (750 mill.) og oppgitt maksverdi (1450 mill.). Halvparten av deltakerne fikk spørsmål om 700 og 1400 mill., den andre halvparten om 800 og 1500 mill., altså en verdi innenfor og en verdi utenfor intervallet mellom P15 og P85.

2.4 Spørsmål til dommergruppen (D-skjema)

Disse fikk den samme innledningen og graf med sannsynlighetsfordeling som de andre gruppene, men ble bare bedt om å besvare to spørsmål. Begge ble gitt i to varianter, A og B.

- DQ1 Setningsutfylling. I variant A skulle de velge hvilket av tallene 750 mill., 1100 mill. og 1450 mill. som er mest naturlig i utsagnet «Det er fullt mulig at det vil koste» I variant B var spørsmålet hvilket av uttrykkene lite sannsynlig, fullt mulig og sannsynlig som er mest naturlig i utsagnet: «Det er at det vil koste 1450 mill.». Vi forventet at A-utsagnet fylles ut med 1450 mill. (i tråd med ekstremitetseffekten), men det var mer usikkert om også B-utsagnet, hvor tallet var oppgitt, ville oppfattes som fullt mulig (fremfor f.eks. lite sannsynlig).
- DQ2 Vurdering av riktighet (1-7). I variant A skulle de vurdere riktigheten av estimatet hvis faktisk kostnad ble (1) 1350 mill. og (2) 1450 mill., altså en kostnad under og en lik med P85. I Variant B skulle vurdere riktighet hvis faktisk kostnad ble (1) 1450 mill. og (2) 1550 mill., dvs. en kostnad lik med P85 og en annen over. Vi tenkte oss at variant B ville gi større forskjell i vurderingen av riktighet, i tråd med resultatene fra Teigen, Løhre og Hohle (2018).

3 Resultater

3.1 Resultater fra hovedskjema (Gruppe 1-3)

Vi beskriver her først anslag og vurderinger av min og maks-verdier, og deretter de verbale karakteristikkene av anslagene.

3.1.1 Anslag av min og maks-verdier (Q1)

Gruppe 1-3 var i Q1 bedt om å anslå P15 og P85, basert på P50 verdien 1100 mill.

Gjennomsnittsanslag er rapportert i Tabell 2 (fem deltagere som anga høyere verdi på min enn maks er fjernet for videre analyser).

Tabell 2. Gjennomsnittsverdier (medianer i parentes) for estimerte min og maks-verdier for veiprosjekt, med oppgitt forventet kostnad 1100 mill., i tre utvalg.

	1: Fagfolk	2: Prosjektkurs NTNU	3: BB-kurs (UiO)
Utfyllingstid	18 min*	11 min	11.5 min
Q 1 KS1 Veiprosjekt (P50 oppgitt til 1100 mill.)			
Estimert Min (P15)	954 mill. (945)	939 mill. (917)	701 mill. (837)
Estimert Maks (P85)	1343 mill. (1300)	1482 mill. (1300)	1640 mill. (1492)
Konfidensintervall (P85-P15)	389 mill. (330)	543 mill. (375)	939 mill. (900)

* Bare deltakerne i denne gruppen besvarte spørsmål Q2 om å formulere usikkerhet med egne ord

Gruppe 1 og 2 angir min-maks konfidensintervaller som er snevrere enn de faktisk anslatte i KS1 (750-1450 = 700 mill.). Dette gjelder 91% av deltagerne i Gruppe 1 og 65% av deltagerne i Gruppe 2, mot bare 47% i Gruppe 3. Enveis ANOVA gir $F(2, 119) = 6.265, p = .003$, Post Hoc tester (LSD) viser at intervallene til Gruppe 3 er signifikant høyere enn intervallene til Gruppe 1 ($p = .001$) og 2 ($p = .014$), mens Gruppe 1 og 2 ikke er signifikant forskjellige fra hverandre. Svarene kan ses på bakgrunn av at BB-studentene nylig hadde diskutert overkonfidens og overestimerte intervaller på kurset.

3.1.2 Sannsynligheter for min og maks-verdier (Q6)

Gruppe 1 var i Q6 spurt om å anslå sannsynligheten for verdier rundt P15, rundt P50, og rundt P85. De gjennomsnittlige sannsynlighetsanslagene, oppgitt i Tabell 3, virker betydelig overdrevne, idet summen blir langt over 100% til tross for at de tre oppgitte verdiene ikke uttømmer alle muligheter som inngår i fordelingen. Sannsynligheten for en høy verdi er angitt som nesten dobbelt så stor som sannsynligheten for en lav verdi, til tross for at deltagerne skal vurdere hva de tror ekspertene mener, etter at de i Q1 selv anslo ekspertenes antatte estimater. Noen

deltakere kan ha misforstått spørsmålet til å gjelde kumulative sannsynligheter. Gruppe 2 og 3 ble gitt et enklere spørsmål som kan besvares direkte ut fra den oppgitte fordelingen, hvor under min = 15%, mellom min og maks = 70% og over maks = 15%. Likevel er det bare halvparten av svarene som summeres til 100%, for de øvrige er summen høyere, spesielt fordi de mener at sannsynligheten for å overstige maks (P85)-verdien er omtrent det dobbelte av sannsynligheten for et resultat under min (P15). Halvparten av deltakerne i Gruppe 2 fikk grafen over fordelingen gjentatt samtidig som de besvarte dette spørsmålet, uten at det så ut til å ha noen betydning for svarene.

Tabell 3. *Gjennomsnittsanslag av sannsynligheter for estimerte kostnader svarende til omtrentlige verdier (± 50 mill., Gruppe 1) og intervaller (Gruppe 2 og 3)*

Gruppe 1: Fagfolk		Gruppe 2: Prosjektkurs NTNU		Gruppe 3: BB-kurs UiO	
		Med graf	Uten graf		
Rundt P15	28.0%	\leq P15	13.2%	10.7%	16,6%
Rundt P50	62.8%	P15-P85	72.4%	83.6%	75.6%
Rundt P85	50.9%	\geq P85	28.9%	25.6%	30.6%
Sum	141.7%		114.5%	119.9%	122.4%

Tabell 4. *Gjennomsnittsverdier for estimerte min og maks-verdier for KS2 og annet veiprojekt i tre utvalg.*

	1: Fagfolk	2: Prosjektkurs NTNU	3: BB-kurs (UiO)
Q7			
Ny forventet verdi	1224 mill.	1249 mill.	1230 mill.
Ny minimumsverdi	1063 mill.	1016 mill.	852 mill.
Ny maksimumsverdi	1439 mill.	1522 mill.	1730 mill.
Nytt konfidensintervall	376 mill.	506 mill.	878 mill.
Q 9 Vurderes som skandaleprosjekt			
hvis sluttkostnad over	1693 mill.	1980 mill.	2365 mill.
Q 10 Annet prosjekt har forventet kostnad 2500 mill.			
Anbefalt budsjett	2830 mill.	2748 mill.	2912 mill.
Q 11 Sannsynlighet < 2500 mill	31.5%	41.6%	42.8%

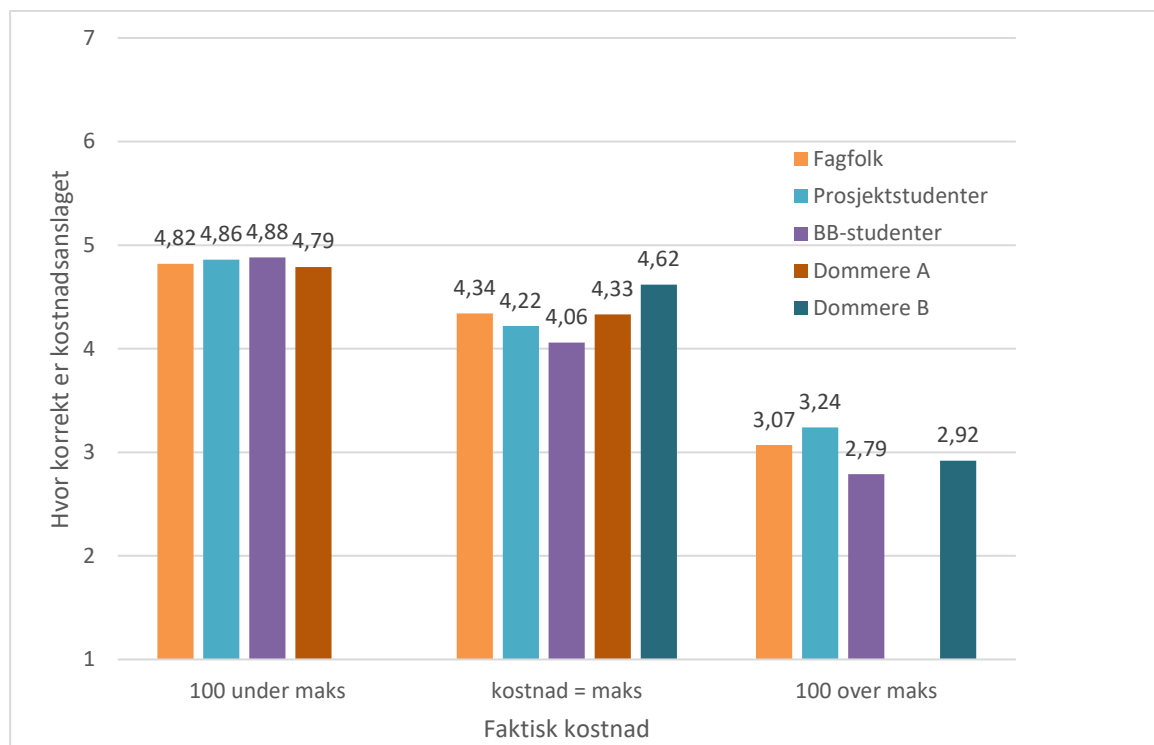
3.1.3 Annet prosjekt (Q10 og Q11)

For et prosjekt med forventet kostnad 2500 mill. kroner, anbefaler et flertall i alle grupper et noe høyere budsjett (2700-3000 mill.). De avviker dermed fra etablert praksis om å bruke P85

estimatet som kostnadsramme. De mener også at det er mindre sannsynlig at kostnadene blir lavere enn at de blir høyere, se Tabell 4 (de to nederste radene).

3.1.4 Vurdering av estimatene (Q8)

Figur 1 viser hvordan bedømt riktighet av estimatene avhenger av faktisk kostnad (de første tre stolpene i hvert sett svarer til gjennomsnittsvurderinger for gruppe 1-3). Når faktiske kostnader tilsvarer maksimums estimatet, ble estimatene vurdert som noe over middels gode av deltagerne i alle tre gruppene ($M = 4.22$). Hvis faktiske kostnader ligger 100 mill. under maksimumsgrensen ble estimatene vurdert som noe bedre ($M = 4.85$). Hvis de derimot ender 100 mill. over maksimum ble estimatene vurdert som klart feilaktige ($M = 3.03$). De to første settene av estimater i Figur 1 er nærmere hverandre ($M_{diff} = 0.63$) enn de to siste ($M_{diff} = 1.19$). En 2 x 3 mixed ANOVA gir signifikant effekt av disse differanseskårene, $F(1, 125) = 6.499$, $p = .012$, men ingen forskjell av gruppe og ingen signifikant interaksjon. Det skjer altså en klar vurdering av estimatet som dårligere når faktisk kostnad passerer den markerte øvre grensen.



Figur 1. Hvor korrekt (1-7) er kostnadsanslaget når faktisk kostnad ligger rundt den estimerte maksverdien? (Over 4 vil si at prediksjonen er mer riktig enn feil, under 4 at den er mer feil enn riktig.)

3.1.5 Verbale sannsynligheter

Forventninger i egne ord (Q2). Dette spørsmålet ble bare besvart av fagfolk i gruppe 1.

Eksempler på svar (usikkerhetsord uthevet her):

- «*Stor sjanse* for at det vil koste opp mot 1400 mill. grunnet *usikkerhet* i grunn. *Liten sjanse* for besparelser».
- «Veien vil *sannsynligvis* koste mellom 1265 og 935 mill., men det er en *stor sjanse* at den vil koste mer eller mindre»
- «*Liten sjanse* for at veien blir billigere enn antatt, *stor sjanse* for betydelig merkostnad»

Kommentar: Det første og det tredje utsagnet viser at ikke alle aksepterer premissene om en symmetrisk fordeling. Det midterste utsagnet illustrerer at et «sannsynlig» kostnadsintervall ikke utelukker at også andre utfall kan ha «stor sjanse» for å forekomme.

Ord som går igjen er *usikkerhet* (28), *sjanse* (23), *sannsynlighet* (20), *mulig* (13). Sjeldnere uttrykk er *forventet* (5), *risiko* (3) *trolig*, *antas*, *potensial*, *overraskelse* og *sikkerhet* (1 hver). (Deltakerne kan ha tatt utgangspunkt i ord som ble nevnt som eksempler i instruksjonen.)

*Sjans*e og *sannsynlighet* opptrer gjerne med adjektiver. *Sjans*e er oftere stor (15) enn liten (7), mens *sannsynlighet* er like ofte stor som liten. *Usikkerhet/ usikkert* er oftest nevnt uten adjektiv (19), når ikke den er «stor». *Mulighet/ mulig* er bare gradert i ett tilfelle. En tredjedel av deltagerne (14) nevner spesifikt usikkerhet i forbindelse med grunnforholdene. Det går igjen at det særlig er overskridelser/fordyrelser deltagerne er opptatt av, til tross for at grafen viser like stor sannsynlighet for høye som for lave kostnader. Videre viser flere eksempler at en stor sannsynlighet for ett utfall ikke utelukker «stor sjanse» for noe annet. Den hyppige henvisning til usikkerhet (i grunnforhold, prisstigning, kompetanse, tid og logistikk) gir inntrykk av en uforutsigbarhet som det kan være vanskelig å sette tall på.

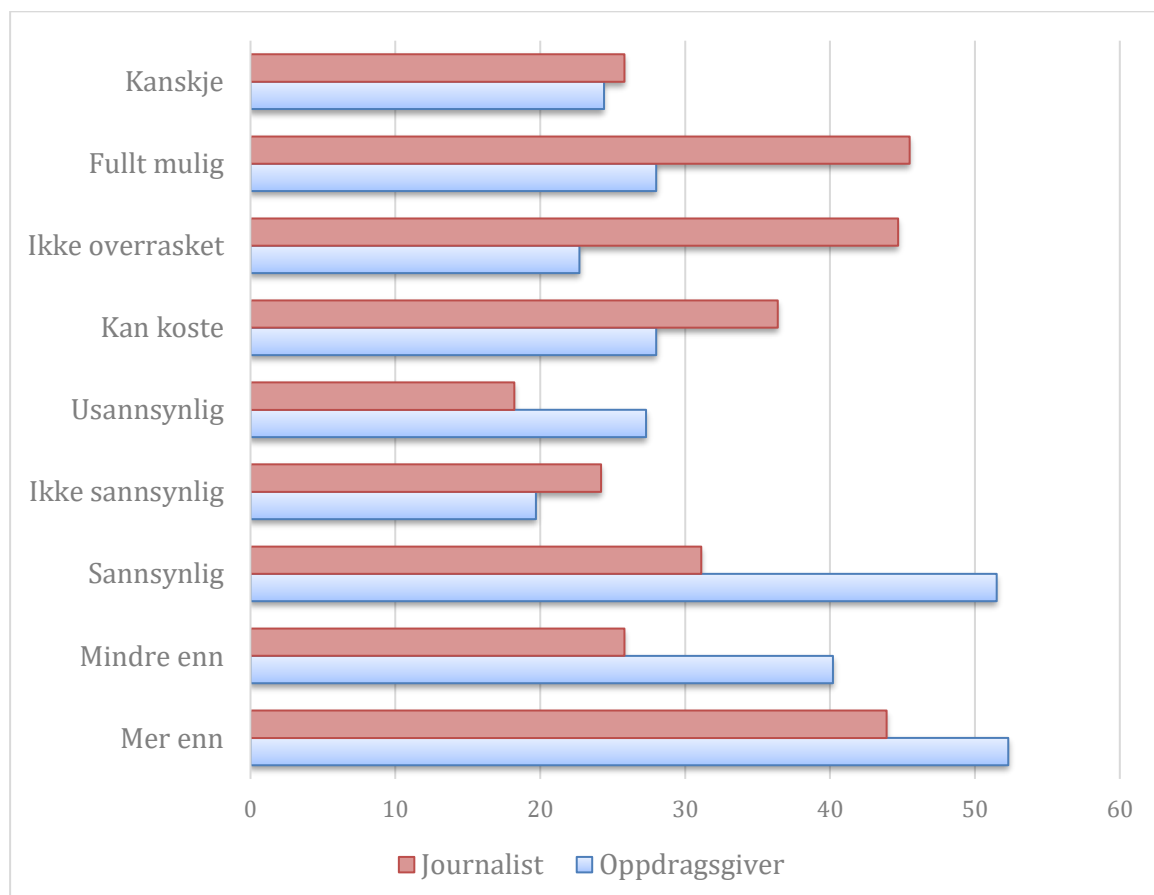
Tabell 5. *Kostnader som passer inn i setninger med ulike verbaluttrykk (medianverdier og prosent av deltagerne som oppgir kostnadsbeløp under antatte minimums- og over antatte maksimumsverdier)*

Uttrykk	Gruppe 1: fagfolk			Gruppe 2: NTNU			Gruppe 3: UiO		
	mn	≤ min (%)	≥ maks (%)	mn	≤ min (%)	≥ maks (%)	mn	≤ min (%)	≥ maks (%)
Mer enn	1000	56.8	9.1	935	56.8	4.5	800	73.5	0.0
Mindre enn	1300	2.3	75.9	1300	6.8	68.2	1800	0.0	85.3
Sannsynlig	1100	0.0	12.2	1100	2.3	9.3	1100	2.9	11.8
Ikke sannsynlig	1200	32.5	45.0	1100	34.9	37.2	1550	26.5	58.8
Usannsynlig	1000	52.4	33.3	1300	23.3	48.8	1125	47.1	44.1
Kan koste	1250	9.5	55.2	1200	6.8	36.4	1200	2.9	26.5
Ikke overrasket	1300	2.4	59.5	1300	2.3	38.6	1350	0.0	51.2
Fullt mulig	1275	2.4	50.0	1257	2.3	27.3	1225	0.0	26.5
Kanskje	1200	2.4	28.6	1200	2.3	27.3	1125	2.9	14.7

Kostnadsbeløp som passer inn i et språklig utsagn (Q3). Deltakere fra alle gruppene fylte ut setninger med passende kostnadsbeløp. Tabell 5 gir en oversikt over hvilke tallverdier som passer sammen med ulike verbaluttrykk i uformelle utsagn om kostnader. I tabellen oppgis medianverdier som blir mindre påvirket av ensidig ekstreme tallverdier (slike forekom spesielt ved uttrykk som «mindre enn», «usannsynlig» og «ikke sannsynlig»). Det går frem av tabellen at *mer enn* svarer til antatte minimumsverdier, *mindre enn* til maks-verdiene, og *sannsynlig* til oppgitt mest sannsynlig (forventet) kostnad (1100 mill.). Ved *Ikke sannsynlig* og *usannsynlig* er medianen lite representativ fordi disse uttrykkene gjerne fullføres med høye eller lave tall som gir en bimodal fordeling: enten over maks eller under min. *Kanskje*, *kan koste*, *ikke overrasket* og *fullt mulig* antas i alle gruppene å bli brukt sammen med høye verdier, i mange tilfeller tilsvarende de anslåtte maksimumsverdier eller mer.

Skandaleprosjekt (Q9). For å snakke om skandaleprosjekt måtte kostnadene komme opp i rundt 2000 mill. i gjennomsnitt (se gjennomsnitt og medianverdier for Gruppe 1-3 i Tabell 4). I alt 95% av deltagerne oppgav «skandaleverdier» tilsvarende de antatte maksimumsverdier eller høyere, $M = 508$ mill. ($Mdn = 392.5$) over anslått maksverdi. Studentgruppene lå her noe høyere enn gruppen fagfolk.

Hvilke uttrykk vil du bruke / hvilke vil en journalist merke seg (Q4-Q5)



Figur 2. Verbaluttrykk som velges brukt til oppdragsgiver/av journalist (prosent av deltagerne i Gruppe 1-3)

Alle tre gruppene foretrakk utsagn med *mer enn*, *sannsynlig* og *mindre enn* i samtale med oppdragsgiver (se Figur 2). De antok derimot at journalister foretrekker mer uformelle uttrykk som *fullt mulig*, *ikke overrasket* og *kan koste*, samt slike som signaliserer at prosjektet kan bli dyrt (*mer enn*). Her viste alle gruppene det samme mønster.

3.2 Resultater forenklet skjema (Gruppe 4)

3.2.1 Verbale sannsynligheter

SQ1 På dette skjemaet var P15 og P85 markert (benevnt min/maks og lave/høye anslag i to varianter av skjemaet). Som ventet beskrev de fleste (rundt 2/3) kostnader mellom 750 og 1450 mill. som sannsynlige, mens kostnader under Mdn = 750 mill. og over Mdn = 1450 mill. ble oppfattet som usannsynlige. Dette gjaldt enten P15 og P85 ble kalt for min/maks eller lave/høye verdier.

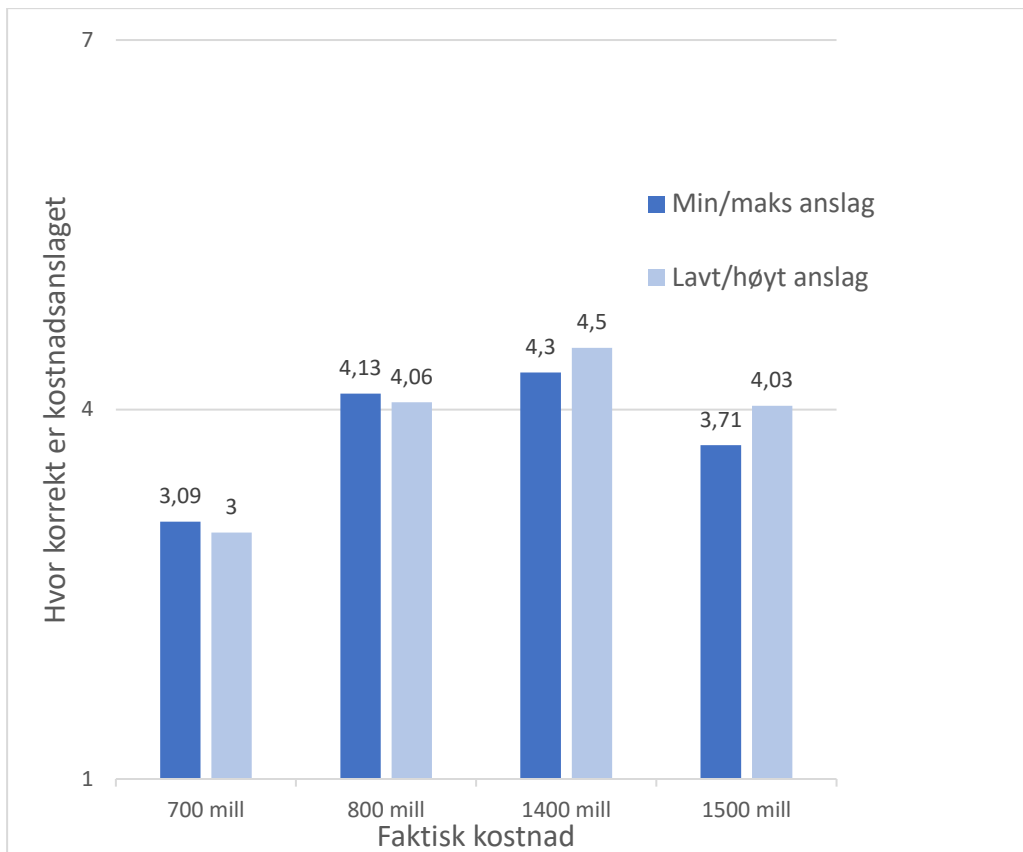
SQ2 Tabell 6 viser at *mer enn*-utsagn typisk fylles ut med den oppgitte verdien for P15, mens *kan koste*, *ikke overrasket* om det koster, og *fullt mulig* at det koster blir fylt ut med høye verdier, gjerne svarende til den oppgitte verdien for P85 (1450 mill.). Medianene er de samme enten P15/P85 kalles minimum/maksimum eller lave/høye verdier.

Tabell 6. *Kostnader som passer inn i setninger med ulike verbaluttrykk, Gruppe 4 (medianverdier og prosent av deltagerne som foreslår lave og høye kostnadsbeløp)*

Uttrykk	Mdn	≤ 750	≥ 1450
Mer enn	750	71.8%	1.5%
Kan koste	1450	5.3%	51.9%
Ikke overrasket	1300	2.3%	43.8%
Fullt mulig	1450	0.8%	59.5%

3.2.2 Vurdering av estimatene

SQ3 Figur 3 viser at anslagene blir vurdert som klart bedre når faktiske kostnader ligger like over enn like under P15, $F(1, 127) = 21.78, p < .0001$, og tilsvarende litt bedre om de ligger under enn over P85, $F(1, 127) = 9.10, p = .04$. Det synes ikke å spille noen vesentlig rolle hvordan det høye eller lave estimatet er benevnt, altså om for eksempel om 750 mill. kalles minimum eller lavt anslag. (Tendensen til at «høyt» anslag er mer korrekt enn «maks» er ikke signifikant.) Det ser imidlertid ut til at anslagene oppfattes som mer korrekte ved høye faktiske kostnader enn ved lave, selv om de bommer/treffer like mye; $M_{1400-1500} = 4.14$ vs. $M_{700-800} = 3.56, t(131) = 3.39, p = .001$.



Figur 3. Hvor korrekt (1-7) er kostnadsanslaget (750-1450) når faktisk kostnad ligger nær de estimerte min- og maksverdiene? (Gruppe 4) (Over 4 vil si at prediksjonen er mer riktig enn feil, under 4 at den er mer feil enn riktig.)

3.3 Resultater dommere (Gruppe 5)

3.3.1 Verbale sannsynligheter (DQ1)

Også dommerne mente at «fullt mulig» pekte mot en høy verdi, fremfor en lav eller en middels (se Tabell 7). Koplingen gikk begge veier, idet de også mente at en høy verdi burde kalles «fullt mulig» fremfor «lite sannsynlig» eller «sannsynlig». Dette skjer til tross for at den høye verdien er definert som P85 i en kumulativ fordeling og dermed har en relativt liten sjanse (rundt 15%) for å inntreffe.

3.3.2 Vurdering av estimatene (DQ2)

Dommere som fikk skjema A vurderte hvor gode estimatene var hvis kostnadene ble 1350 mill. (under maks) og 1450 mill. (lik maks). Gjennomsnittsvurderingene er inkludert i Figur 1 (4. stolpe) og viser at estimatet blir best med kostnad på 1350 mill., men at det er liten forskjell på vurderingene i disse to tilfellene, $M_{diff} = 0.46$. De som fikk skjema B (5. stolpe) vurderte forskjellen mellom 1450 mill. (lik maks) og 1550 mill. (over maks) som atskillig større, $M_{diff} = 1.70$; $t(246) = 10.99$, $p < .001$. Estimatet blir altså vurdert som langt mer feil idet kostnadene passerer maks-verdien.

Tabell 7. Hvilket kostnadsbeløp er «fullt mulig», og hvilket uttrykk passer til å beskrive maksverdien i en usikkerhetsfordeling? Utfylt av to dommergrupper.

<i>Utsagn</i>	<i>Alternativ</i>	<i>Prosent valg</i>
Variant A (N = 196)		
Det er fullt mulig at det vil koste	750 mill. (P15)	6.1%
	1100 mill. (P50)	14.3%
	1450 mill. (P85)	79.6%
Variant B (N = 160)		
Det er at det vil koste 1450 mill.	lite sannsynlig	16.9%
	fullt mulig	68.1%
	sannsynlig	15.0%

4 Diskusjon

Vi stilte innledningsvis spørsmål om hvordan fagfolk (og legfolk) (1) vurderer størrelse og sannsynligheter knyttet til usikkerhetsintervallet, (2) hva de oppfatter som korrekte anslag, og (3) hvordan de bruker ord som uttrykker sannsynlighet og usikkerhet.

4.1 Snevre intervaller og forventninger om overskridelser

Fagfolk oppga min.- og maks.-verdier som gjorde at nesten alle (91%) usikkerhetsintervallene deres ble for snevre, omtrent halvparten av det som var angitt i KS1. Bare BB-studentene oppga intervaller av tilnærmet riktig størrelse, noe som antakelig bør ses på bakgrunn av at underestimering av intervaller var gjennomgått på kurset.

Alle utvalgene oppga min.- og maks.-verdier som var noenlunde symmetriske rundt «forventet» verdi, men av de øvrige svarene gikk det frem at de faktisk trodde overskridelser var mer sannsynlige enn kostnader under det forventede. Dette fremgår av fagfolkenes åpne svar på spørsmål Q2, hvor mange poengterte at det var store sjanser for fordyring, mens innsparelsen ble ansett som lite sannsynlige. Det fremgår også av de høye sannsynlighetene de oppga for et resultat rundt P85. Siden noen deltagere i denne gruppen kan ha misforstått spørsmålet, fikk Gruppe 2 og 3 et enklere spørsmål, nemlig om sannsynligheten for kostnader over P85. Denne sannsynligheten anslo de fortsatt som større enn sannsynligheten for kostnader under P15, selv om spørsmålet gjaldt hva de trodde den opprinnelige ekspertgruppen ville svare. Endelig tenkte de at estimatene i KS2 ville ligge ca. 100 mill. høyere enn ved KS1.

Forventninger om fordyrelser og overskridelser kan ha utspring i erfaringen. Det er en kjent sak at mange store offentlige prosjekter er blitt betydelig mer omfattende, har tatt lenger tid og blitt atskillig mer kostnadskrevenne enn opprinnelig antatt (Flyvbjerg, 2016; Morris & Hough, 1987). De opprinnelige anslagene kan dels vært satt for lavt av strategiske grunner, dels har man ikke tatt høyde for uventede og som regel fordyrende avvik fra de opprinnelige forutsetningene. I tillegg får overskridelser betydelig mer oppmerksomhet i media enn prosjekter som holder seg innenfor rammene. En gjennomgang av Concepts KS2-prosjekter (Welde, 2017) viser en mer likelig fordeling av faktiske kostnader over og under den forventede verdi, mens våre deltagere ikke venter at sjansen for fordyrelser og besparelser vil være noenlunde like, til tross for at de fikk oppgitt en symmetrisk sannsynlighetsfordeling. Imidlertid stemmer det med deres vurderinger at det gjerne finner sted en kostnadsutvikling mellom KS1 og KS2.

4.2 Vurdering av estimerer: en grenseeffekt

I Gruppe 1-3 ble estimatene oppfattet som klart dårligere når faktiske kostnader oversteg maksimumsgrensen enn når de lå på antatt maksimum, som vist i figur 1, selv om overskridelsene var relativt små (100 mill. kroner, dvs. 7-8% av medianestimatet). Om de ligger tilsvarende under maksimumsgrensen spiller mindre rolle. Her ser vi altså en effekt av å markere P15 og P85 som grenseverdier for et anslag, selv om det gikk tydelig frem av sannsynlighetsfordelingen at man også regner med at høyere og lavere verdier kan forekomme. Om man hadde valgt å markere f.

eks. P10 og P90 i stedet (med tilsvarende utvidet intervall mellom grenseverdiene) er det tenkelig at større faktiske kostnader ikke lenger ville oppfattes som overskridelser og ville bli bedømt som mer korrekt anslått.

Den bedømte forskjellen mellom verdier innenfor og utenfor intervallet er likevel mindre enn i tidligere studier vedr. klimarelaterte prognoser. Teigen, Løhre og Hohle (2018) fant at grensene for intervallet var viktigere enn både mest sannsynlig verdi og grad av sikkerhet når det gjaldt å bedømme om en intervallprognose var korrekt eller ikke. Deltagerne i disse studiene bedømte bare ett utfall hver og hadde altså ikke muligheter til å sammenligne, slik som deltakerne i Gruppe 1-3. Flomprognosene var heller ikke illustrert med en graf som viste sannsynlighetsfordelingen. I tillegg var grensene for intervallet oppgitt av ekspertene selv, mens deltakerne i Gruppe 1-3 var henvist til verdier de selv hadde foreslått, og som de altså ikke kunne vite svarte til ekspertintervallene. Dermed ble det også mindre klare grenseverdier, og kanskje mindre grunn til å oppfatte intervallene som kategorier.

Faktisk blir P85 for vedtatte prosjekter gjerne definert som *kostnadsrammen*, og danner slik en reell grense for hvilke fordyringer som kan tolereres og hvilke som ikke oppfattes som akseptable. Dermed er våre deltagere i noen grad på linje med retningslinjene for bruk av estimer innen KS-ordningen. Vi har imidlertid liten grunn til å tro at våre respondenter (uten kanskje enkelte i gruppen «fagfolk») hadde noe kjennskap til denne praksis. Effekten av definerte grenser var spesielt tydelig i dommergruppen, som vel må regnes som eksperter på skjønsmessig beslutningstaking, men neppe på KS-ordningen for store statlige prosjekter.

I Gruppe 4 (Figur 3) var kostnadenes avvik fra maksimums- og minimumsverdier oppgitt til å være svært små (pluss minus 50 mill. kroner, dvs. 3-4 prosent). Dermed kan estimatene utenfor og innenfor intervallet kan ha blitt oppfattet som «praktisk talb» på grensen til P15 og P85. I tillegg var grafen gjentatt på samme side som spørsmålet. Den gjør det lett å se at estimatet beskriver en kontinuerlig fordeling, og legger forholdene slik til rette for et mindre skarpt skille mellom «riktige» og «gale» estimer. Det er litt overraskende at kostnader rundt minimumsgrensen blir oppfattet som mye dårligere anslått enn kostnader rundt maksimum. Dette er imidlertid forenlig med en uttalt forventning om at kostnader oftere blir høye enn lave. Erfaringen (og logikken) tilsier at det er en nedre grense for hvor billig noe kan bli, mens det ikke fins en tilsvarende øvre grense. Lave kostnader blir dermed mer overraskende, til tross for at fordelingen er tegnet symmetrisk. En slik tolkning forutsetter at deltagerne ikke har full tiltro til estimatet slik det fremstilles i grafen.

4.3 Verbale sannsynligheter: en ekstremitetseffekt

I stedet for å be deltakerne tilkjenne hvilke tallmessige sannsynligheter som svarer til ulike verbale uttrykk (oversettelsesmetoden / «the translation approach») har vi i denne undersøkelsen villet kartlegge hvordan ulike verbale uttrykk brukes i praksis til å karakterisere ulike utfall i en kjent sannsynlighetsfordeling (utfyllingsmetoden / «the which outcome approach»; Teigen, Juanchich & Filkuková, 2014). Tidligere studier har vist at ord og uttrykk som er forenlig med utfall av varierende sannsynlighet (*kan, mulig*) oftere brukes om høye enn om middels og lave verdier (extremity effect). En ekstremitetseffekt har også blitt påvist for ord som uttrykker lave sannsynligheter (*usannsynlig*). I sistnevnte tilfelle siktes det gjerne til verdier som ikke har

forekommet eller man antar ikke vil forekomme, og som dermed har mye lavere sannsynlighet enn de som avdekkes ved oversettelsesmetoden (hvor «unlikely» gjerne oppfattes som svarende til sannsynligheter rundt 20%; Theil, 2002).

Kan koste og fullt mulig viste i alle gruppene en ekstremitetseffekt, ved at de stort sett passet best sammen med høye verdier, gjerne tilsvarende P85-verdiene eller mer. Verdiene ble særlig høye i Gruppe 4 (Tabell 6) og Gruppe 5 (Tabell 7) som fikk høye oppgitte maksverdier de kunne ta utgangspunkt i. Interessant nok fant vi en lignende effekt for *ikke overrasket over*, som man kanskje skulle tro ville henviser til sannsynlige verdier, men som altså passet best til høye verdier som pr. definisjon skulle være lite sannsynlige. Deltagerne synes å være enige om at de som sier vi er «ikke overrasket» uttaler seg om kostnader det kunne være en viss grunn til å bli overrasket av. Omtaler av høye verdier med slike «positive» uttrykk kan ha uheldige konsekvenser, ved at en utenforstående lett vil komme til å tro at slike utfall også er sannsynlige (Teigen et al., 2018).

Usannsynlige og ikke sannsynlige verdier er også ekstreme, men var i denne undersøkelsen ikke begrenset til svært høye verdier. I tidligere studier (Jenkins et al., 2018; Teigen et al., 2013) handlet en «usannsynlig» temperatur eller en «usannsynlig» stigning av havnivået typisk om temperaturer eller havnivåer som var usannsynlig høye. Det fremgår av tabell 5 at usannsynlig og ikke sannsynlig i en kontekst av kostnader like ofte viser til kostnader som er usannsynlig lave. En mulig tolkning er at tendensen til å rapportere ekstremt høye verdier i dette tilfelle motvirkes av en generell forventning om at prosjekter ofte blir dyrere enn antatt. En økning på 50% eller tilmed en fordobling av kostnadene virker dermed mindre usannsynlig enn en halvering av dem.

4.4 Begrensninger

Resultatene er basert på svar på hypotetiske spørsmål om et enkelt veiprojekt, slik det ble vurdert av en uensartet samling grupper med begrenset ekspertise. En annen type prosjekt som deltagerne kanskje kjente bedre kunne gitt rom for andre tolkninger og vurderinger. I gruppen fagfolk, som spørsmålene vel var mest relevant for, var svarprosenten lav fordi skjemaet var distribuert i form av frivillig åpning av lenke (vi kan likevel håpe at de som besvarte skjemaet hørte til de over middels interesserte). Til gjengjeld fikk vi god dekning i gruppen dommere (høy svarprosent og mange deltagere), men denne studien ga bare rom for to utvalgte spørsmål.

Vi ser likevel et mønster i besvarelsene som går igjen fra gruppe til gruppe, fra studenter uten fagkunnskap, studenter som er mer fortrolig med prosjekter og beslutningsforskning, til fagpersoner som gjennom sin yrkeserfaring har måttet diskutere og vurdere kostnader og prognoser for tilsvarende prosjekter. Disse fellesdragene kan oppsummeres med blant annet disse stikkord: Overpresisjon for intervaller, tendens til overvurdering av sannsynligheter, bruk av min og maks-verdier for å avgjøre riktigheten av et estimat, og størst fokus på høye verdier i omtale av kostnader. På det siste punktet har våre deltagere og de faktiske beslutningstakere en felles interesse, idet det er de høye verdiene som definerer kostnadsrammen og bestemmer om en kostnad kan betraktes som akseptabel (jfr. debatten om hvorvidt forsvarets kampfly-prosjekt har «sprukket» eller ikke; Johansen & Landmark, 2016).

Vi har i alle skjemaene gjort usikkerheten mest mulig forståelig og tilgjengelig ved å illustrere den med en klokkeformet sannsynlighetsfordeling, som vi mener er mest 'pedagogisk' når det gjelder

å få frem betydningen av P15 og P85 verdier som vilkårlige punkter i en kontinuerlig fordeling. Her kunne det vært på sin plass med en studie som sammenligner denne fremstillingen av usikkerhetsfordelingen med en S-formet kumulativ fordeling (som er vanlig i KS-rapportene), eller uten figur, med bare oppgitte ytterpunkter for et intervall (som i Teigen et al., 2018). Vi vet heller ikke hvordan riktighetsvurderingene ville bli hvis markeringene for P15 og P85 var fjernet. En hypotese ville være at alle avvik fra P50 da ville bedømmes som dårligere anslått, fordi de ikke lenger befinner seg innenfor et definert prediksjonsintervall.

Riktighetsvurderingene i Figur 1 kan være påvirket av at de samme deltagerne blir bedt om å ta stilling til flere ulike utfall, og nødvendigvis vil sammenligne dem med hverandre. Et between-Ss design (slik som i Teigen et al., 2018) ville her vært mer realistisk idet samme prosjekt vanligvis bare har ett utfall. Slike design vil kreve et større antall deltagere, og kunne vanskelig gjennomføres med de utvalgene som ble rekruttert til denne undersøkelsen. Endelig ville det ha styrket prosjektet om vi hadde lyktes i å rekruttere et utvalg faktiske beslutningstakere (i departement eller stortingskomiteene) som er målgruppen for KS-rapportene.

5 Konklusjon

Det er ikke selvsagt at en leser av en KS-rapport vil oppfatte P15, P50 og P85-verdiene i en sannsynlighetsfordeling etter boka. Sannsynlighetsanslagene i pilotstudiene og hovedstudien var vanskelige å tolke, og tyder på at misforståelser lett kan oppstå selv der definisjonene er klare. Hvis man skal være sikker på at de oppfattes som intendert, er det antagelig ikke nok å gi en enkel formell definisjon og overlate implikasjonene til en leser av rapporten.

Betegnelse *minimum*, *maksimum* og *forventet* (eller *mest sannsynlig*) verdi forenkler budskapet, men kan også villedes ved å oppfattes for bokstavelig eller kategorisk. En leser kan forledes til å sette likhetstegn mellom «mest sannsynlig» og sannsynlig, eller anta at estimatet er feil hvis faktisk kostnad overstiger maksimum (selv om estimatet innebærer at dette skjer fra tid til annen). Det kan videre synes paradoksalt at mange lesere vil anslå usikkerhetsintervallet til å være snevrere enn det som oppgis i KS-rapporten, samtidig som de fleste tror budsjettoverskridelser (les: undervurderinger av kostnader) hører med til dagens orden. Det virker altså som om potensielle brukere har en tendens til å forestille seg at estimatene er mindre dristige, men også mer feilaktige enn det som faktisk er tilfellet. Også den språklige omtale av kostnadsanslag er preget av et fokus på høye verdier, idet tilsynelatende nøytrale betegnelser om hva en antar *kan* skje, er *fullt mulig* eller som man *ikke vil bli overrasket over* antas å bli brukt om høye fremfor middels eller lave estimater, med fare for at slike estimater blir oppfattet som mer sannsynlige enn beregnet.

Lesernes unøyaktighet i forståelse av estimater innebærer også en utfordring for estimeringsprosedyren. Et usikkerhetsintervall er klart mer dekkende enn enkle punktestimater, og en sannsynlighetsfordeling enda bedre. Samtidig kan det være et problem at ikke alle former for usikkerhet lar seg like lett bake inn i den samme modellen. Det har vært hevdet at slike formelle projeksjoner ikke er like gode til å ta vare på «unknown unknowns», det vil si kilder til usikkerhet man ikke kjente til eksistensen av og derfor ikke heller ikke kunne gi et anslag av på forhånd. Det kan likevel gis åpning for at noen områder er mer preget av slike «genuine overraskelser» enn andre, for eksempel på grunn av høy kompleksitet, nyhet eller forutsetninger under endring (Parker & Risbey, 2018). Fagfolkene i vår undersøkelse fremhevet spesielt den store *usikkerheten* (f. eks. usikre grunnforhold ved veibygging) og syntes her å sikte til et moment som kom *i tillegg* til de estimerte kostnadene og kunne gjøre prosjektet mer omfattende og mer krevende enn anslått.

Vår studie indikerer at selv velutdannede lesergrupper i blant kan trenge ytterligere «voksenopplæring» for å forstå probabilistiske kostnadsestimater. Leverandørene av slike estimater kunne på sin side ha nytte av å kjenne noen hovedspor i lekfolks tankeverden, og derved øke sjansen til å rydde misforståelser av veien der de lettest oppstår. Det er for eksempel ikke sikkert at man kommer leseren i møte ved å snakke om minimum, maksimum og forventede estimater, selv om de fleste vel synes det klinger bra. Det som klinger bra klinger ikke alltid rett. Så kanskje en mer utstrakt bruk av tekniske uttrykk som P15, P85 og P50 tross alt er bedre, nettopp fordi de viser til en hel fordeling, og selv en overflatisk leser vil skjønne det er hjelpebegreper som trenger en nærmere forklaring. En visuell fremstilling av fordelingen, som vi

har brukt i denne studien, kan også være et skritt på vei til økt forståelighet, men løser ikke alle vansker.

Denne studien kan følges opp på ulike måter. Vi har ikke lykket med å få (tilstrekkelig) data fra ønskede aktører så datagrunnlaget kunne gjerne vært utvidet. Det kunne også vært gjennomført analyser rundt f.eks. om det er forskjell på usikkerhetsspennet hos etatene og KS-rådgiverne og om de omtaler estimatene annerledes. Vi har også vært forsiktige med å utforme anbefalinger basert på funnene i studien, dette kunne gjerne Concept eller Finansdepartementet sett nærmere på.

Referanser

- Barnes, A. (2015). Making intelligence analysis more Intelligent: Using numeric probabilities. *Intelligence and National Security* 4527: 1–18. <https://doi.org/10.1080/02684527.2014.994955>.
- Berry, D. C., Raynor, D. K., Knapp, P. R., & Bersellini, E. (2003). Patient understanding of risk: impact of EU guidelines and other risk scales for consumer medicines information. *Drug Safety*, 26, 1-11.
- Budescu, D. V., & Wallsten, T. S. (1995). Processing linguistic probabilities: General principles and empirical evidence. *Psychology of Learning and Motivation*, 32, 275-318.
- Dhami, M. K., Mandel, D. R., Mellers, B. A., & Tetlock, P. E. (2015). Improving intelligence analysis with decision science. *Perspectives on Psychological Science*.
- Dieckman, N. F., Peters, E., & Gregory, R. (2015). At home on the range? Lay interpretations of numerical uncertainty ranges. *Risk Analysis*, 35(7), 1281–1295.
- Dovre Group AS & Transportøkonomisk institutt (2009). *E18 Knapstad – E6 i Follo: Kvalitetsikring av konseptvalg (KS1). Hovedrapport*
- EEC (1998). *A guideline on the readability of the label and package leaflet of medicinal products for human use*. Brussels: EC Pharmaceuticals Committee; 1998.
- Finansdepartementet (2008). *Kostnadsestimering. Veileder nr. 6*. Oslo
- Flyvbjerg, B. (2016). The fallacy of beneficial ignorance: A test of Hirschman's hiding hand. *World Development*, 84, 176-189.
- Hacking, I. (2001). *An introduction to probability and inductive logic*. Cambridge: Cambridge University Press.
- Hesselberg, J.-O. (2018). *Dommerseminaret 2018: Spørsmål fra psykolog Jan-Ole Hesselberg*. Oslo: Survey-Monkey
- Hohle, S. M. & Teigen, K. H. (2018). When probabilities change: Perceptions and implications of trends in uncertain climate forecasts. *Journal of Risk Research*. doi: 10.1080/13669877.2018.1459801
- Honda, H., & Yamagishi, K. (2017). Communicative functions of directional verbal probabilities: Speaker's choice, listener's inference, and reference points. *The Quarterly Journal of Experimental Psychology*, 70, 2141-2158.
- Jenkins, S. C., Harris, A. J. L., & Lark, R. M. (2018). Understanding 'unlikely (20% likelihood)' or '20% likelihood (unlikely)' outcomes: The robustness of the extremity effect. *Journal of Behavioral Decision Making*. DOI: 10.1002/bdm.2072

- Johansen, A. & Landmark, A. D. (2016). Kampflyanslag til besvær. *Dagens Næringsliv*, 10. november, s. 42.
- Joslyn, S., Nemeč, L., & Savelli, S. (2013). The benefits and challenges of predictive interval forecasts and verification graphics for end users. *Weather, Climate, and Society*, 5(2), 133-147.
- Kalinowski, P., Lai, J. & Cumming, G. (2018). A cross-sectional analysis of students' intuitions when interpreting CIs. *Frontiers of Psychology*, 9, 112. doi: 10.3389/fpsyg.2018.00112
- Langnickel, F., & Zeisberger, S. (2016). Do we measure overconfidence? A closer look at the interval production task. *Journal of Economic Behavior & Organization*, 128, 121-133.
- Löhre, E. & Teigen, K. H. (2017). Probabilities associated with precise and vague forecasts. *Journal of Behavioral Decision Making*, 30, 1014-1026.
- Mastrandrea, M.D. et al. (2010) Guidance note for lead authors of the IPCC Fifth Assessment Report on consistent treatment of uncertainties. *IPCC*. <http://www.ipcc.ch>
- Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review*, 115(2), 502-517.
- Moore, D. A., Tenney, E. R., & Haran, U. (2016). Overprecision in judgment. In G. Wu & G. Keren (Eds.), *Handbook of Judgment and Decision Making* (pp. 182-209) New York: Wiley
- Morris, P. W. G. & Hough, G. H. (1987). *The anatomy of major projects: A study of the reality of project management*. New York: Wiley.
- Parker, W. S., & Risbey, J. S. (2018). False precision, surprise and improved uncertainty assessment. *Philosophical Transactions of the Royal Society A*, 373, 2014.0453.
- Redelmeier, D., Koehler, D. J., Liberman, V., & Tversky, A. (1995). Probability judgment in medicine: Discounting unspecified possibilities. *Medical Decision Making*, 15, 227-230.
- Riege, A. H. & Teigen, K. H. (2013). Additivity neglect in probability estimates: Effects of numeracy and response format. *Organizational Behavior and Human Decision Processes*, 121, 41-52
- Teigen, K. H. (1988). When are low-probability events judged to be "probable"? Effects of outcome-set characteristics on verbal probability judgments. *Acta Psychologica*, 67, 157-174.
- Teigen, K. H. & Brun, W. (1999). The directionality of verbal probability expressions: Effects on decisions, predictions, and probabilistic reasoning. *Organizational Behavior and Human Decision Processes*, 80, 155-190.
- Teigen, K. H., Filkuková, P., & Hohle, S. M. (2018). It can become 5 °C warmer: The extremity effect in climate forecasts. *Journal of Experimental Psychology: Applied*, 24, 3-17.

- Teigen, K. H., Juanchich, M., & Filkuková, P. (2014). Verbal probabilities: An alternative approach. *Quarterly Journal of Experimental Psychology*, *67*, 124-146.
- Teigen, K. H., Juanchich, M., & Riege, A. H. (2013). Improbable outcomes: Infrequent or extraordinary? *Cognition*, *127*, 119-139.
- Teigen, K. H., & Jørgensen, M. (2005). When 90% confidence intervals are only 50% certain: On the credibility of credible intervals. *Applied Cognitive Psychology*, *19*, 455-475.
- Teigen, K. H., Løhre, E., & Hohle, S. M. (2018). The boundary effect: Perceived post hoc accuracy of prediction intervals. Revised manuscript under review (Judgment and Decision Making)
- Theil, M. (2002). The role of translations of verbal into numerical probability expressions in risk management: A meta-analysis. *Journal of Risk Research*, *5*, 177–186.
- Tversky, A., & Koehler, D. J. (1994). Support theory: A nonextensional representation of subjective probability. *Psychological Review*, *101*, 547–567.
- Ülkümen, G., Fox, C. R., & Malle, B. F. (2016). Two dimensions of subjective uncertainty: Clues from natural language. *Journal of Experimental Psychology: General*, *145*, 1280–1297.
- Welde, M. (2014). Avvik mellom etatenes kostnadsestimat og anbefalingene i KS2. *Concept arbeidsrapport*. Trondheim: Norges teknisk-naturvitenskapelige universitet.
- Welde, M. (2017). Kostnadskontroll i store statlige investeringer underlagt ordningen med ekstern kvalitetssikring. *Concept rapport nr 51*. Trondheim: Norges teknisk-naturvitenskapelige universitet.