

Sammendrag:

Evalueringstradisjoner og kvalitetskriterier i samferdselssektoren

Formålet med denne rapporten har vært å kartlegge kvalitetskriterier som tas i bruk ved ulike typer evalueringer, hvilke likheter og forskjeller vi finner på tvers av evalueringstradisjoner og -tema, og hvilke særegne utfordringer og dilemmaer en står overfor når en skal realisere disse kriteriene.

Vi definerer evaluering som en *systematisk etterundersøkelse av om effektene av offentlig virksomhet er i samsvar med intensjonene, med det formål at slike undersøkelser skal ha relevans for pågående eller fremtidige beslutningsprosesser.*

Når en setter effekter i sentrum for analysen, må en forholde seg til kontrafaktisitetetsproblemet som det avgjørende spørsmålet ved evaluering, dvs. spørsmålet om effektene ville vært annerledes dersom bestemte tiltak ikke hadde vært introdusert.

Fire perspektiver

Vi analyserer evaluering med bakgrunn i fire perspektiver. Det første perspektivet som forstår evaluering som evalueringsmetodikk, har sin bakgrunn i Elviks forståelse av at det er metoden som skiller en god og en dårlig evaluering – ikke resultatet. Avgjørende er det om evalueringen tilfredsstillende kravene til statistisk, teoretisk, intern og ekstern validitet.

Det andre perspektivet, det informerte oppdragsgiverperspektivet, forstår evaluering som en målrasjonell prosess. Evaluering tas i bruk når saksforholdet er vesentlig, dvs. har stor økonomisk eller prinsipiell betydning. Evalueringsdesignet bør fortrinnsvis være formalisert for å kunne sammenligne ulike evalueringer, og det bør være basert på indikatorer som er gode styringsparametre.

Det tredje perspektivet, det statsvitenskapelig informerte meta-perspektivet, slik vi finner det hos Vedung, er pluralistisk. Det bygger på en antagelse om at en kan benytte ulike perspektiver til ulike evalueringsformål, samtidig som han avgrenser evaluering til etteranalyser.

Det fjerde perspektivet, det profesjons- og kunnskapssosiologiske perspektivet, innebærer den forståelse at teorier "begriper" likeartede fenomener på ulike måter. Forskere "finner ikke" empirien, de skaper den også gjennom de måtene teoriene begriper fenomenene. Dette gjør at validitetsspørsmål i noen grad er teori-interne, ulike funn og vurderinger kan være resultater av ulike teorier og modeller, noe som gjør sammenlignbarhet vanskelig og interdisiplinære kontroverser utbredt.

Videre er ulike begreper ”verdilatet”, slik at bestemte teorier gir støtte til noen hensyn og verdier og svekker andre.

To av perspektivene - det informerte oppdragsgiverperspektivet og Vedungs metaperspektiv – er her benyttet avgrensende: De definerer hvilke typer evalueringer vi inkluderer i denne rapporten, og hvilke vi ekskluderer. Denne funksjonen har, i all hovedsak, også Elviks perspektiv om evaluering som evalueringsmetodikk. Vi tar utgangspunkt i at metoden, ikke funnene, bestemmer om evalueringen er god eller dårlig. Det profesjons- og kunnskapssosiologiske perspektivet introduseres for å få fram at et meta-perspektiv må åpne for spørsmålet om sammenlignbarhet mellom perspektiver på ulike fenomener.

Klassifikasjonsskjema for sammenligning av evalueringer

Disse perspektivene gir grunnlag for følgende klassifikasjonsskjema for sammenligning og vurdering av evalueringer:

- Hva er *temaet*, *problemstillingen* eller *formålet* med evalueringen?
- Hva kjennetegner *metode* og *datagrunnlag* (kvalitative eller kvantitative data, eventuelt type underkategori)? Hvor omfattende er undersøkelsen, hvor mange variabler og typer variabler har den?
- Hva kjennetegner *teorien* som blir benyttet, hvilke *forutsetninger* bygger den på, hvilke *antagelser* søker den å undersøke, hva er de sentrale *begrepene*, har den en klar *disiplinforankring*, hva er dens *gyldighetsområde*?
- Er mulige/sannsynlige utilsiktede effekter godt dekket?
- Kumulativitet: Hvordan antas evalueringen å kunne føre til ny kunnskap (substansielt/metodisk/teoretisk)?

I klassifikasjonsskjemaet inkluderes også *kvalitetskriterier*: Hvilke kvalitetskriterier tar de ulike evalueringene i bruk? Har de et design som i utgangspunktet kan tilfredsstillende krav til:

- *Statistisk validitet* (dvs. reliabilitet, representativitet, analyseteknikker og statistiske sammenhenger),
- *Teoretisk validitet* (eksplisitt teorigrunnlag, operasjonalisering, testbare hypoteser, gir funnene støtte til teorien),
- *Intern validitet* (om det i tillegg til den statistiske sammenhengen også kan fastslå årsaksretning, og om denne holder for kontroll med andre forklaringer)
- *Ekstern validitet* (om en kan generalisere resultatene av undersøkelsen utover den spesifikke konteksten, i tid, rom og eventuelt til andre saksområder)

Anvendelse av skjemaet

I rapporten anvender og prøver vi ut dette klassifikasjonsskjemaet på noen utvalgte evalueringer innen samferdselssektoren, primært med sikte på å få fram hypoteser om det spesifikke ved bestemte typer evalueringer eller det vi mer løselig kan omtale som evalueringstradisjoner. Vi skiller her mellom tre typer: Multidimensjonal variabelanalyse, nyttekostnadsanalyse og organisasjonsanalyse. De fleste av de evalueringene som blir gjennomgått har til felles at de har pågått over lengre tid med relativt omfattende ressurser. Det er derfor ikke grunn til å anta at de på det området skiller seg negativt ut fra arbeidssituasjonen ved evaluering generelt, slik at kvaliteten ikke burde være lavere enn i andre tilfeller.

Multidimensjonal variabelanalyse

Den første evalueringen vi gjennomgår og vurderer, er evalueringen av tiltakspakkene for kollektivtransport (kapittel 2). Formålet med denne typen evaluering antar vi at er representativ for det vi har gitt benevnelsen multidimensjonal variabelanalyse: Den omhandler og analyserer en rekke ulike variabler på ulike analysenivå som både omhandler tiltak, rammebetingelser, holdninger og handlinger blant målgruppen (i dette tilfellet passasjerer). Gitt at en, som i dette tilfellet, har diakrone data, er dette et design som i utgangspunktet kan tilfredsstillende kravene til statistisk validitet. I denne konkrete evalueringen finner vi klare svakheter ved selve analysen av data. Dels gjelder dette operasjonell validitet, problemer med å benytte opplevd endret tilfredshet med tilbudet og opplevde endrede reisevaner som indikatorer for tilbudsending og endrede reisevaner. Svakheter ved operasjonelle variabler, kombinert med manglende data for passasjerstatistikk, kan også antas å svekke de aggregerte analysene. Det er et problem med presentasjonen av data-analysen at slike forhold i liten grad drøftes og presenteres på en måte som gjør grunnlaget for regresjonsanalysen transparent. Dermed er det også et paradoks at resultater presenteres med høy grad av presisjon.

De teoretiske begrepene, som ”holdninger”, ”preferanser”, ”faktisk atferd”, er operasjonaliserbare, og i liten grad forankret i en bestemt teoretisk disiplin. Den teoretiske modellen er relativt omfattende, med lite klart formulerte hypoteser. Funn gir derfor ikke grunnlag for å diskutere selve modellens relevans, og det blir heller ikke gjort. Manglende disiplinforankring gjør at ekstern validitet ikke kan drøftes i forhold til bestemte, mer overgripende teorier. I prinsippet kunne en ha diskutert ekstern validitet i betydningen om funnene kan reproduseres på samme type fenomen (tiltak for økt kollektivandel i byområder). Heller ikke dette blir gjort her.

Designet innebærer at evalueringen potensielt kan være kumulativ på to måter - metodisk og substansielt. Metodisk ved oppbygging av databaser som gir grunnlag for komparative statistiske analyser over tid, men dette forutsetter at databasen forbedres og utvikles. Substansielt ved økende grad av kunnskap om faktorer som påvirker kollektivandeler i by. Et hovedfunn er at tiltakene ikke har vesentlig effekt på transportmiddelfordelingen, men i noen tilfeller gir passasjervekst for kollektivtransporten. Også på dette området drøfter forfatterne i liten grad den eksterne validiteten til de konklusjonene som trekkes, som at bedret

frekvens er den viktigste forklaringen til endret reisehyppighet, og at dårligere tilbud gir større passasjerbortfall enn forbedringer gir passasjervekst.

Ex post-nyttekostnadsanalyser

I kapittel tre drøftes to hovedtilnærminger til analyser av feilbudsjettering og kalkyle i større samferdselsinvesteringer. De to tilnærmingene gir grunnlag for en diskusjon om tverrfaglige kontroverser. Den ene typen evaluering som gjennomgås, nyttekostnads-etteranalyser, innebærer en antagelse om at en viktig årsak til feilkalkyler er metodiske svakheter. Etteranalyser ble forsøkt gjennomført for flere jernbaneprosjekter, parallelt med vurdering av tidligere og nåværende metodehåndbøker. Dokumentasjonen for tidligere prosjekter var imidlertid så dårlig at selve etteranalysen bare hadde illustrativ verdi. Derimot ga gjennomgangen grunnlag for konkrete forslag til forbedringer i metodehåndboka.

Formålet med nytte-kostnadsanalyse som etteranalyse er nettopp å forbedre ex ante-metodikken gjennom å: 1) Kartlegge avvik mellom de beregnede samfunnsøkonomiske virkningene i forundersøkelsen og etterundersøkelsen og 2) å lokalisere hvilke av disse som skyldes ulik metode, ytre forutsetninger, prosjektavhengige avvik.

Denne typen analyse er disiplinforankret i velferdsøkonomisk teori, der formålet er å undersøke om et tiltak gir et positivt eller negativt samfunnsøkonomisk resultat. Dette skjer bla ved at en sammenstiller virkninger som så langt mulig tilstrebes gitt en felles pengeverdi. Utgangspunktet for slike analyser er begrepene om konsument- og produsentoverskudd. Et prosjekt hvor den samlede neddiskonterte nytten (målt i kroneverdi) er større enn de samlede neddiskonterte kostnader (også i kroneverdi) kan sies å være samfunnsøkonomisk lønnsomt og representerer følgelig akseptabel ressursbruk for samfunnet.

En nytte-/kostnadsanalyse skal vise tiltakets effekt i forhold til ulike aktører i samfunnet. Her skilles det gjerne mellom noen hovedkategorier av interessenter: Det offentlige, brukerne og samfunnet for øvrig. I samferdselssektoren framkommer ofte effektene for det offentlige i form av investeringskostnader, endringer i drifts- og vedlikeholdskostnader og offentlig kjøp av transporttjenester. Brukerne er både operatører og trafikanter, som kan oppleve endret tidsbruk (reisetid, ventetid, gangtid) og endring i direkte økonomiske størrelser som driftskostnader og utlegg (som for eksempel billettkostnader). Samfunnet for øvrig uttrykker det som også kalles eksterne effekter, som for eksempel endringer i lokale, regionale og globale utslipp, støy og naturinngrep.

Det teoretiske rammeverket er preget av at antagelser og forutsetninger er eksplisitt og klart formulert. Undersøkelsen tilfredsstillende kravene til et klart teorigrunnlag ved at begreper er operasjonaliserbare. Det mest kritiske punktet i operasjonaliseringsprosessen er usikkerhet ved operasjonalisering av komponenter som i utgangspunktet ikke er beskrevet i økonomiske termer: Tidskomponenter og eksterne kostnader som miljø- og ulykkeskostnader. Teorien gjør det også mulig å teste antatte sammenhenger mellom sentrale variabler, men (selvsagt) ikke den teoretiske forutsetningen at dette er de relevante variablene. I sum tilfredsstillende denne typen evaluering godt kravene til teoretisk validitet.

Når det gjelder statistisk validitet, opererer denne type analyse med to typer statistiske sammenhenger, for det første de sammenhenger som uttrykkes gjennom selve nytte-kostnadsanalysen, som det krever spesialisert kunnskap for å vurdere og er lite transparent, og for det andre framstillingen av avvikstabeller som gir en oversiktlig avviksanalyse.

Selve det metodiske opplegget er designet for å ivareta de fleste sidene ved intern validitet: Systematisk gjennomtenkning av årsaks-virkningsforhold. Ekstern validitet og generaliserbarhet, er primært knyttet til utvikling av metodikken, dels som gjennomgang og forbedring av selve metodehåndbøkene og forbedret og mer ensartet metodikk for veg og jernbane. Det er også i form av utvikling av metodeverktøy at denne typen undersøkelser er kumulativ. Oppsummert er dette en evalueringstradisjon preget av høy grad av formalisering, som imøtekommer de mest grunnleggende krav Elvik har formulert til evalueringsmetodikk. Formålet er å utvikle og forbedre metodeverktøyet slik at en gjør færre feil i framtidige nytte-kostnadsanalyser.

Men det er særlig én side ved prosjektgevinster og -kostnader som ikke blir empirisk etterprøvd gjennom et slikt opplegg, og det er usikkerheten knyttet til konvertering av tids-, miljø- og ulykkeskostnader til økonomiske kategorier. For andre typer variabler kan ex post nyttekostnadsanalyse sees på som et forsøk på å lokalisere metodiske svakheter som kan medføre feilbudsjettering og -prioritering innen samferdselssektoren.

Kritikken som er reist mot denne metoden (dvs. mot nyttekostnadsanalyse per se og ikke primært mot etteranalyser) er knyttet til forutsetninger som gjøres og som ikke testes. Dette gjelder særlig spørsmålet om tidsverdsetting og verdsetting av kostnader knyttet til ulykker, støy osv. Slike problemstillinger er kontinuerlig diskutert og analysert innenfor den typen fagmiljø som jobber med nyttekostnadsanalyse. Typisk i denne sammenheng vil være å basere verdsettingen på individenes egne verdsettinger, eventuelt utvidet til kartlegging av ulike trafikantsegmenters verdsettinger. Kritikken som reises utenfra, fra forskere i andre fagtradisjoner, er at selve de forutsetningene som gjøres for slike analyser er problematiske. I rapporten gjengis Sverre Strands kritikk, der han peker på at tidsverdien gjerne vil være den største verdien i mange nyttekostnadsanalyser av samferdselsprosjekter, og at det er visse grunnlagsproblemstillinger knyttet til verdsetting av tid. Han skiller mellom tre trinn i tidsanalyser: 1) Måling av tid og tidsbesparelser, 2) aggregering av tidsbesparelser og 3) pengeverdering av tidsbesparelser. Spørsmål 1 omhandler tema som hvor mye større enn null en enkeltbesparelse må være for at den skal kunne omtales som signifikant nok til at tiden kan anvendes alternativt. Problem 2 omhandler under hvilke forutsetninger tidsbesparelser kan aggregeres, mens det tredje punktet handler om konvertering av tidsverdier til pekuniære verdier. Strand hevder at alle tre er problematiske, men at standardtilnærmingen til problemet vil være at en gjennom tidsverdsetting løser 1 og 2.

Tilsvarende rettes det systematisk kritikk mot nytte-kostnadsanalyse for en rekke andre antagelser som den bygger på, som aktørforutsetninger innen konsumentteori som konverterer det kritikere oppfatter som moralske, kollektive problemstillinger til spørsmålet om individuelle preferanser, og problemer knyttet til bruk av neddiskonteringsraten på miljøspørsmål. Denne fageksterne kritikken

kan sees på som ledd i en interdisiplinær kritikk, som hevder at det å benytte denne type analyse medfører å fremme bestemte verdier. Kritikerne peker også på problemet med manglende transparens ved slike modeller. Rapporten drøfter i noen grad kritikken av kritikken, men understreker at dette mest av alt viser behovet for offentlig, saklig tverrfaglig debatt.

Organisasjonsevalueringer (I)

Den andre tradisjonen for evaluering av overskridelser og feilbudsjettering i samferdselsinvesteringer som drøftes i denne rapporten, er Aalborg-miljøets forskning på feltet. I tillegg til å peke på metodiske forklaringer som årsak til slik feilbudsjettering, analyserer en i denne tradisjonen også organisatoriske og institusjonelle variabler. Aalborg-miljøets forskning blir i første rekke presentert gjennom Bengt Flyvbjergs arbeider. Flyvbjerg forstår overskridelser som resultat av desinformasjon og lokaliserer tre hovedmodeller for å forklare slik desinformasjon: Tekniske modeller som refererer til feil i data og metoder, psykologiske modeller som refererer til kognitive feil og politisk-økonomiske modeller i form av strategisk misrepresentasjon og løgn. Han argumenterer for at de to første modellene ikke kan forklare desinformasjon og feilbudsjettering, blant annet fordi en da skulle forvente læring og redusert feilbudsjettering over tid. Fordi de andre forklaringsmodellene ikke finner støtte gjenstår den siste. Den beste forklaringen blir da at insitamentstrukturen er slik at fordelene med å desinformere er større enn ulempene og resultatet blir "survival of the unfittest" – det er ikke det beste prosjektet som vinner, men det som har best desinformasjon. Botemidlene for dette problemet er dels utvikling av bedre metoder for å estimere kostnader, nytte og risiko gjennom "reference class forecasting", dels endrede insitamentstrukturer som vil begrense desinformasjon og fremme riktig informasjon.

Teoretisk er tilnærmingen til Flyvbjerg eklektisk uten en klar disiplinforankring. Fordelen med en slik tilnærming er at den spenner opp et relativt bredt analytisk rammeverk som inkluderer mange relevante variabler - både metodiske, psykologiske og organisatoriske. På den andre siden er aktørforutsetningene og antagelsene om institusjoners virkemåte noe underspesifiserte, men drøftingen indikerer at disse er forankret i institusjonell økonomi. Det kritiske punktet i analysen er den teoretiske og interne validiteten. Det er for eksempel få klare empiriske indikatorer på løgn, og denne forklaringsvariabelens kraft fastlegges gjennom et kontrafaktisk resonnement. Det metodiske problemet Flyvbjerg står overfor er det klassiske for de som forsker innen tradisjoner som omfattes av mistankens hermeneutikk, det er vanskelig å finne empirisk støtte for at løgnere lyver, enten de vet det selv eller ikke.

I rapporten argumenterer vi for at det analytiske inntaket til Flyvbjerg – "organisation matters" – representerer en fruktbar dreining og et viktig supplement til utvikling av metodeverktøyet. Vi argumenterer imidlertid også for at viktige elementer i selve tolkingen av disse organisatoriske trekkene og aktørforutsetningene, er problematiske. En mulig retning for forbedring av denne forskningen vil være å introduserer et mer finmasket variabelsett både når det gjelder organisasjonsvariabler og trekk ved aktørene. Forskningsprosjekter med dette temaet kunne gi mer kunnskap om bakgrunnen for overskridelser og slik ha

praktisk-politisk betydning for hvordan vi organiserer beslutningsprosessen, samt være et bidrag i den internasjonale faglige debatten.

Organisasjonsevalueringer (II)

I det siste kapitlet drøftes organisasjonsevalueringer mer systematisk. Kapitlet innledes med en evaluering der betingelsene for en systematisk undersøkelse av effekter av organisasjonsendring lå godt til rette, fordi en både hadde den tid, de ressurser og det datamateriale som tillot kartlegging av endring både på effektvariablene (trivsel og sykefravær) og på tiltaksvariabel (introduisering av team). Kapitlet fortsetter med en gjennomgang av en organisasjonsanalyse der en ikke hadde effektvariabler: Evalueringen av TP 10 som prosess. Denne tar for seg forsøk på samordnet areal- og transportpolitikk i ti større norske byområder og har status som en prosessevaluering. Den teoretiske modellen for evalueringen er lite formalisert, men avgrenser seg til å presentere en analytisk modell som beskriver sentrale aktører, prosjektdesign, mål for og organisering av lokale prosesser, og endelige produkter i form av planer og vedtak. Effektene av slike vedtak måles ikke, blant annet fordi de antas å være av så vidt langsiktig karakter.

Det analytiske skjemaet impliserer ikke spesifisering av hvilke variabler som kan antas å påvirke utfallet. Dataene baserer seg også her på flere typer metoder; intervjuer, seminarer, spørreskjema. Basisen for tolkningen er mer omfattende beskrivelser der noen sentrale begreper (som sentrale aktører, føringer) har heuristisk karakter og strukturerer drøftingen. Analysen presenterer ikke et eksplisitt teoretisk grunnlag, men presenterer noen begreper og antatte sammenhenger forankret i statsvitenskap og organisasjonsteori. Designet bygger ikke på statistisk analyse og kan dermed heller ikke sies å tilfredsstille kravene til statistisk validitet.

Manglende eksplisitt teoretisk grunnlag gjør det vanskelig både å vurdere den interne og den eksterne validiteten. Designet er slik lite formalisert og vurdering av evalueringen må basere seg på vurdering av de substansielle tolkingene til forskere og andre. Gyldighetsområdet og den eksterne validiteten avgrenser seg dermed til samordnet areal og transportpolitikk i norske byområder. På dette feltet har den imidlertid også hatt effekter. En strøm av utredninger om svakheter ved organiseringen av den lokale areal- og transportpolitikken fulgte i kjølvannet av denne evalueringen. Den har altså vært kumulativ i den forstand at den har vært oppfattet å representere sentrale kjennetegn ved sektoren.

Generelle utfordringer

For det første er det et fellestrekk med de evalueringene som har vært gjennomgått i denne rapporten, at de i liten grad drøfter ekstern validitet, og dermed i liten grad andre undersøkelser om samme emne. Det mest klare unntaket er nytte-kostnadsanalysen, der utvikling av modellverktøyet er sentralt. For det andre er det også, med samme unntak, et fellestrekk at undersøkelsene i liten grad eksplisitt knytter seg til overordnede teoritradisjoner. Begge deler innebærer at det er liten grunn til å anta at slike evalueringer vil bli underlagt den kvalitetskontrollen og muligheten for nyansering og kunnskapsutvikling som følger av systematisk komparasjon og teoretisk diskusjon. Det er dermed

avgjørende at en enten legger forholdene til rette for slik videreutvikling, eller at en etablerer funksjonelle ekvivalenter til den type kritikk som ligger i den offentlige debatt.

Videre avdekker rapporten to typer transparensproblemer. Det transparensproblemet som følger med komplekse statistiske analyser, som øker når datagrunnlaget ikke er offentlig, innebærer at utenforstående vanskelig kan få innsikt i fenomenet. Avgjørende er her den kontrollen som ligger i kollegial kritikk. Denne type transparensproblem kan assosieres til det Elster har kalt "hard obskurantisme" – selv om dette begrepet også inkluderer at kunnskapen i tillegg til å være lite transparent også er lite relevant. Det andre transparensproblemet finner vi i kvalitative evalueringer der modeller er lite formaliserte, slik at antagelser og resonnementer vil være vanskelig tilgjengelige. På dette området er den valideringen som ligger i møtet med informantene og tilbakemeldinger et sentralt kontrolelement. Disse to ytterpunktene representerer også ulike måter der evaluerer kan påvirke resultatet, bevisst eller ubevisst. I statistiske analyser kan dette skje gjennom rekoding- og modelleringsprosessen, i kvalitative analyser i sammenfatningen og tolkingen av data. Til forskjell fra kvantitative analyser, vil en organisasjonsanalyse uten effektvariable, ikke kunne håndtere kontrafaktisitetproblemet.

Vår rapport viser hvordan de hovedtypene av kvalitetskriterier som Elvik stiller gjennom sin analyse, i all hovedsak er sentrale og fruktbare redskaper for å vurdere evalueringer, og å få fram kritiske punkter ved ulike evalueringstradisjoner. Å stille krav om teoretisk, intern og ekstern validitet medvirker til å skjerpe blikket for styrker og svakheter ved evalueringer, og å etablere en felles målestokk. Derimot er det gode argumenter for å fravike kravene til statistisk validitet, fordi dette i noen tilfeller vil innebære å forkaste muligheten for å få fram kunnskap som er vesentlig, men ikke kan skaffes gjennom statistiske analyser.

Implikasjonene av det profesjons- og kunnskapsteoretiske perspektivet kommer fram i drøftingen av evalueringer av større samferdselsinvesteringer. For det første viser to ulike tilnærminger, nyttekostnadsetteranalyser og organisasjonsanalyser, hvordan ulike tilnærminger kan fremskaffe komplementær kunnskap. For det andre er kritikken og diskusjonen om nyttekostnadsanalyser et eksempel på interdisiplinære kontroverser som bla omhandler hvordan modellene kan fremme bestemte verdier.

Profesjonsperspektivet benyttes også til å sammenfatte trekk ved koblinger mellom evaluerere, oppdragsgivere og grad av transparens. Følgende hypotese formuleres: Epistemiske samfunn som er organiserte rundt spesialisert kunnskap som rommer lite transparente modeller, kan representere bestemte problem – for omgivelsene. Manglende metodiske overveielser og kritikk kan, i de tilfeller det epistemiske samfunn er organisert rundt komplekse statistiske analyser, medføre entydige og sikre vurderinger på spørsmål som er flertydige og sammensatte – som rangordning av prosjekter i nyttekostnadsanalyser. I andre tilfeller der slike epistemiske samfunn er organisert rundt bestemte politiske problemer og har liten affinitet til alternativanvendelse av ressurser, vil manglende metodiske overveielser gi støtte til en ekspanderende policy-sektor. Logikken i resonnementet vil være at når dette tiltaket ikke virket, så øker vi medisinen,

snarere enn å diskutere 0-hypotesen: At andre forhold er avgjørende. Mens evalueringen i mange tilfeller burde ha ledet til politisk nøkternhet og kanskje til og med resignasjon, blir den en satsplanke for fornyet optimisme.

Den antatte relevansen av en slik hypotese styrkes av at det er et fellestrekk for alle de evalueringene vi har gjennomgått, at de i en viss forstand er teoriimmune. De er preget av at en ikke undersøker om teorien som ligger til grunn for undersøkelsen støttes av resultatet av undersøkelsen. Dette innebærer at det er mulig og sannsynlig at det etableres epistemiske samfunn på tvers av evaluering og forskning, som baserer seg på "common beliefs" som ikke lar seg avkrefte. Dette øker betydningen av tverrfaglig og tverrsektoriell kommunikasjon og kritikk, og dermed betydningen av institusjonelle ordninger som medvirker til dette.