

Technical University of Denmark



## **Fra teori til praksis - om evaluering af arbejdsmiljøindsatser i en ustyrlig virkelighed** Baggrundsnotat til workshop 3 om metoder til måling og evaluering af arbejdsmiljøindsatser

**Hasle, Peter; Nielsen, Klaus T.; Limborg, Hans Jørgen ; Seim, Rikke; Bramming, Pia**

*Publication date:*  
2012

[Link back to DTU Orbit](#)

### *Citation (APA):*

Hasle, P., Nielsen, K. T., Limborg, H. J., Seim, R., & Bramming, P. (2012). Fra teori til praksis - om evaluering af arbejdsmiljøindsatser i en ustyrlig virkelighed: Baggrundsnotat til workshop 3 om metoder til måling og evaluering af arbejdsmiljøindsatser. Center for forskning i virkemidler og arbejdsmiljøindsatser, CAVI.

## **DTU Library** Technical Information Center of Denmark

---

### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# Fra teori til praksis

## – om evaluering af arbejdsmiljøindsatser i en ustyrlig virkelighed

---

### Baggrundsnotat til workshop 3 om metoder til måling og evaluering af arbejdsmiljøindsatser

Peter Hasle, Klaus T. Nielsen, Hans Jørgen Limborg, Rikke Seim og Pia Bramming  
Center for forskning i virkemidler og arbejdsmiljøindsatser (CAVI)

---

## 1. Resumé

Dette notat fokuserer på, hvordan evaluering af arbejdsmiljøindsatser i praksis kan designes og anvendes til forbedring af fremtidige indsatser. Evaluering er en stærk, generel tendens på samfundsplan, hvilket naturligvis også gør sig gældende på arbejdsmiljøområdet. Som redegjort for i det første notat, er arbejdsmiljøområdet imidlertid præget af såkaldte 'vilde problemer' med komplekse årsagssammenhænge, hvilket kan gøre det vanskeligt at gennemføre gode evalueringer af arbejdsmiljøindsatser. Desuden er det meget omkostningsfuldt at gennemføre evalueringer. Ofte ender evalueringer derfor i, hvad man kan kalde 'pragmatisk evaluering', der kan være en kombination af et eller flere af de evalueringsparadigmer – responsiv evaluering, monitorering, effektevaluering og realistisk evaluering – som blev præsenteret i det første notat. Valget af evalueringsdesign vil altid være begrundet i de tilgængelige ressourcer og den politiske virkelighed, desuden vil valget være meget afhængig af hvilken type arbejdsmiljøindsats, der er tale om – en begrænset indsats med et relativt lille budget, en omfattende indsats, en tværgående indsats eller et større program med flere koordinerede aktiviteter. I notatet gennemgås evalueringsscenarier for disse fire typer indsatser.

Det er i stigende grad et krav, at alle beslutninger om fremtidige arbejdsmiljøindsatser skal baseres på evidens for således at sikre, at indsatsen fører til det ønskede resultat. Evidens kan defineres som information, der giver en stærk begrundelse for noget. Men derudover er evidens også generelt blevet en institutionaliseret praksis for, hvordan viden og evaluering bruges til at udvikle politikker og indsatser. Evidens anskues ofte ud fra to parametre – validitet og effekt. Validitet – eller gyldighed – deles traditionelt op i ekstern og intern validitet, hvor den interne validitet handler om den anvendte metodes gyldighed, fokuserer den eksterne validitet på mulighederne for at generalisere resultatet til en anden sammenhæng. Effekt kan opdeles i output, outcome og impact. Output er det umiddelbare, synlige produkt af en indsats, outcome er de umiddelbare virkninger indsatsen har over for målgruppen, mens impact er den mere langsigtede virkning. Af pragmatiske grunde evalueres ofte på outputs af indsatser, fx antal downloads af en given BAR-vejledning. Disse outputs må dog ikke forveksles med opnåelse af det ønskede resultat af BAR-vejledningen som fx nedbringelse af en specifik form for arbejdsulykke.

I notatet beskrives fire eksempler på, hvordan et evalueringsdesign mest fornuftigt kan tilpasses de tilgængelige ressourcer. Det drejer sig om: *Den begrænsede indsats*, hvor udfordringen er at finde den simplest mulige evalueringstilgang, der kan bidrage med afgørende viden. *Den omfattende indsats*, der kræver klare programteorier, for at kunne prioritere evalueringer. *Den tværgående evaluering*, der særligt kræver en afklaring af målgrupper og relevante evalueringstyper for de enkelte målgrupper. Endelig er der *programmet*, hvor evalueringen må bygges op således, at det bliver muligt at gennemføre tværgående analyser, hvilket igen kræver refleksioner over programteori.

## 2. Indledning

Formålet med dette tredje notat er at undersøge, hvordan diskussionen af evaluering af arbejdsmiljøindsatser kan gå fra teori til praktisk anvendelighed. I det første notat blev evaluering defineret som en systematisk vurdering af aktiviteter med henblik på at få grundlag for at træffe fremtidige beslutninger. Det medfører, at evaluering som begreb lægger op til en praktisk anvendelse. En evaluering skal hjælpe beslutningstagere til at træffe fremtidige beslutninger. Det betyder imidlertid ikke, at der ikke også kan være andre formål, fx at legitimere fortidige eller fremtidige beslutninger, at forhale ubehagelige beslutninger eller genere politiske modstandere. Men vi skal i det følgende – måske lidt troskyldigt – forudsætte, at de evalueringer vi diskuterer, hovedsageligt har som ambition at forbedre kvaliteten af fremtidige beslutninger.

Evaluering som et både praktisk og teoretisk felt, har en historie på et halvt hundrede år, og evaluering har udviklet sig fra at være enkeltstående, relativt sjældne aktiviteter til at være et gennemgribende element i nutidens samfund. Således har evaluering fået et sådan omfang, at nogen decideret taler om 'evalueringssamfundet' (Dahler-Larsen 2003). Nu igangsætter beslutningstagere – specielt i den offentlige sektor – sjældent aktiviteter uden de på den ene eller anden måde evalueres, og uden at aktiviteterne i forlængelse heraf kan begrundes med evidens fra tidligere evalueringer, at de leverer de resultater som forventes. Denne udvikling præger også arbejdsmiljøområdet, hvor der også stilles stadig større krav til evaluering. Det gælder ikke mindst i 2020-planen, hvor en række punkter indeholder krav om evaluering (se notatet til workshop 1).

Arbejdsmiljøområdet er dog – som så mange andre områder – i den situation, at der sjældent er simple kausale sammenhænge mellem en konkret indsats og den ønskede effekt på sikkerhed og sundhed. Dels er årsagsvirkningskæden lang med mange muligheder for brud undervejs, og dels kan mange arbejdsmiljøproblemer og de tilknyttede indsatser betegnes som vilde problemer, hvor komplekse sammenhænge gør det umuligt at være sikker på sammenhængen mellem indsats og effekt. Et simpelt eksempel er afskærmning af rundsave. Som konkret løsning er det ganske lige til: En skærm forhindrer, at fingrene får kontakt med savklingen. Men spørgsmålet om implementeringen af indsatsen kan i høj grad forstås som et *vildt* problem. Kravet om afskærmning er en af de ældste arbejdsmiljøregler, men branchen kæmper stadig med at få den overholdt, fordi der i praksis er en række forhold omkring økonomi, produktivitet, viden, kultur, professionsidentitet og meget andet, som betyder, at der alligevel mangler skærme på mange rundsave.

Der er mange forskellige tilgange til evaluering, og indledningsvist pegede vi på fire dominerende paradigmer. Ingen af disse kan vælges som det *rigtige* paradigme. De har hver deres styrker og svagheder og kan derfor være velegnede til at opfylde bestemte evalueringsformål (Tabel 1).

Paradigme	Styrker	Svagheder
Responsiv (proces-) evaluering	<ul style="list-style-type: none"> <li>• mulighed for at forbedre indsatsen undervejs</li> </ul>	<ul style="list-style-type: none"> <li>• begrænset viden om resultatet af indsatsen</li> </ul>
Monitorering	<ul style="list-style-type: none"> <li>• mulighed for at følge en langsigtet udvikling og dermed reagere på den</li> </ul>	<ul style="list-style-type: none"> <li>• begrænset viden om resultatet af implementeringsaktiviteterne</li> </ul>
Effektevaluering	<ul style="list-style-type: none"> <li>• viden om resultatet af en indsats</li> </ul>	<ul style="list-style-type: none"> <li>• begrænset viden om årsagerne til det fremkomne resultat</li> </ul>
Realistisk (virknings-) evaluering	<ul style="list-style-type: none"> <li>• viden om de mekanismer som i forskellige kontekster fører til resultater</li> </ul>	<ul style="list-style-type: none"> <li>• vanskeligt at generalisere mekanismer og deres relation til konteksten</li> </ul>

**Tabel 1: Styrker og svagheder ved de fire evalueringsparadigmer**

Problemet er, at det er omkostningsfuldt at gennemføre gode og gyldige evalueringer, og at beslutningstagerne ofte vil ønske at få resultater, som bedst opnås gennem evalueringer, der indeholder elementer af flere eller alle paradigmerne. Ofte ender evalueringer derfor i, hvad man kan kalde et femte paradigme – pragmatisk evaluering. Det kan der være gode begrundelser for, men der er også en oplagt risiko for at sætte sig mellem flere stole. Den information, som kommer ud af en for sammensat evaluering med ønsker om lidt fra alle hylder, kan være så usikker at den reelt ikke bidrager til at forbedret beslutningsgrundlag.

Det er denne problemstilling, som er hovedtemaet for dette notat. Vi fortsætter i næste afsnit med en diskussion af evidens, som er det begreb, der anvendes om viden i tilknytning til evaluering og politiske beslutninger. Dernæst diskuterer vi problemstillingen i forbindelse med overførsel af viden fra evaluering til beslutninger, og vi slutter notatet med at pege på forskellige evalueringsbehov afhængigt af arbejdsmiljøindsatsernes karakter. Der er åbenlys forskel på, om der er tale om evaluering af små projekter, store projekter eller egentlige programmer, og om specifikke eller mere tværgående evalueringer. Ligesom der er stor forskel på, om man ønsker viden om, hvordan en indsats er blevet udbredt, om den har nået målgruppen, om den har ført til konkrete ændringer på arbejdspladsen eller om den reelt har reduceret omfanget af sygdomme og skader.

### 3. Evidens

Evidens er blevet et modebegreb og dermed også et begreb, som bruges og misbruges til mange forskellige formål. Det er i stigende grad et krav, at alle beslutninger om nye aktiviteter skal baseres på evidens for at sikre aktiviteterne fører til de ønskede resultater. Nye forslag kan derfor skydes ned med henvisning til, at der ikke foreligger evidens, eller diskussionen kan drejes til at være en teknisk diskussion af gyldigheden af

den foreliggende evidens. Håndteringen af dette spørgsmål vender vi tilbage til lidt senere. Indledningsvis undersøger vi her evidensbegrebet lidt nærmere.

Ifølge Nudansk Ordbog<sup>1</sup> betyder evidens: 'En oplysning som støtter en antagelse' eller med Oxford Advanced Learner's Dictionary<sup>2</sup> formulering (oversat til dansk af os): 'Information som giver en stærk begrundelse for at tro noget eller beviser noget'. Ifølge denne betydning omfatter evidens altså enhver oplysning, som understøtter en antagelse. Evidens behøver således ikke at være resultatet af forskning eller videnskab. Men evidens – oplysninger – giver også anledning til at rejse kritiske spørgsmål: Er oplysningen nu relevant for den pågældende antagelse? Og kan man overhovedet stole på oplysningen? Når man anvender begrebet evidens, indebærer det derfor altid en diskussion af relevans og gyldighed.

Et eksempel herpå kan være asbestsagen i 1980'erne. Der blev anvendt flere milliarder kroner på at fjerne loftsplader i dagsinstitutioner og skoler, men den sundhedsmæssige betydning heraf kan i høj grad diskuteres. En væsentlig årsag til at bruge de mange penge var en udtalelse af en læge i fjernsynet: Én asbestfiber er tilstrækkeligt til at fremkalde kræft. Mens denne oplysning ud fra en biologisk betragtning må antages at være korrekt, er den næppe relevant for vurderingen af risikoen ved henholdsvis at fjerne loftspladerne eller ved at forsegle og bevare loftspladerne.

Men ud over at være information, der giver en stærk begrundelse for noget, er evidens også blevet en del af en institutionaliseret praksis for, hvordan viden og evalueringer på samfundsplan bruges til at udvikle politik og indsatser, "så vi er sikre på, at de virker". Som nævnt i første notat så taler man om Cochrane-tilgangen inden for det medicinske felt og Campbell-tilgangen inden for de bredere samfundsmæssige områder, som arbejdsmiljøområdet er et eksempel på. Inspirationen til tænkning om evidens-baseret politik kommer fra det medicinske område. Generelt står forestillinger om meget skarpe metodekrav til, hvad der overhovedet tæller som evidens stærkt i både Cochrane- og Campbell-tilgangene. Begge tilgange favoriserer de såkaldte Randomiserede Control Trials (RCTs), hvor det er tilfældigt (randomiseret), hvem der modtager behandlingen eller indsatsen, og hvor effekten måles både hos modtagerne af behandlingen/indsatsen og hos en kontrolgruppe, som ikke modtog den. Begge tilgange ser tilsvarende bort fra den information, som stammer fra kilder, der ikke lever op til sådanne skrappe metodekrav, hvorfor det ofte kan være et problem, at der ikke er meget evidens at basere sin politik på, når den gyldige viden skal sammenvejes. Desuden forsøger man gennem disse tilgange at se bort fra betydningen af kontekst, og det er derfor vanskeligt at opnå viden om, i hvilke sammenhænge evidensen faktisk er relevant.

Andre forståelser giver forslag til, hvordan man kan arbejde med evidens, der er tættere koblet til praksis. Et oplagt eksempel er Pawsons kritisk realistiske tilgang til evidensbaseret politik (Pawson 2006), som vi allerede har refereret flittigt til. Mere bredspektret har man, fx på uddannelsesområdet, set en række forsøg på at omfavne, og måske samtidig også opbløde, en evidensbaseret tilgang til udviklingen af politikker og praksis (fx Hargreaves 1997, Davies m.fl. 2000, se også Simons 2003). Endeligt kan man pege på Rieper & Foss Hansen (2007), der problematiserer den evidensstærke tilgang, som placerer RCT-produceret evidens i toppen af hierarkiet. Deres bud er, at det er mere relevant at tale om evidensstypologier frem for

---

<sup>1</sup> Politikens Nudansk Ordbog med etymologi. København, 2000.

<sup>2</sup> Oxford Advanced Learner's Dictionary. Oxford 1995. Opslagsordet er *evidence*.

om et evidenshierarki, for så vidt som forskellige typer af metoder producerer viden, som har forskellig relevans afhængigt af den sammenhæng, den producerede viden skal bruges i.

### **3.1. Validitet**

Gyldighed af en oplysning, eller validiteten som det kaldes i forskning, er som nævnt det ene af to kriterier for evidens. Validiteten deles traditionelt op i to typer (Krogstrup 2011): intern og ekstern validitet. Den interne validitet handler om den anvendte metode (teoretisk og praktisk) faktisk er i stand til at påvise de sammenhænge, som man vil undersøge. I epidemiologiske undersøgelser er et væsentligt spørgsmål for den interne validitet således at kontrollere for alle alternative årsager til den årsags-virkningssammenhæng som undersøges. Mens den eksterne validitet handler om mulighederne for at generalisere resultatet til en bredere sammenhæng. Det vil sige om den effekt, som er fundet i en undersøgelse, også kan forventes at opstå i en anden sammenhæng – i en anden kontekst.

Et eksempel fra epidemiologisk forskning i det psykosociale arbejdsmiljø kan illustrere denne problemstilling. Det er en nærmest fastslået kendsgerning, at høj indflydelse på arbejdet har en positiv indflydelse på helbredet. Det er påvist i mange undersøgelser og opfattes på mange måder som en relativt kontekstfri sammenhæng. En nylig finsk undersøgelse af skovarbejdere (Joensuu et al. 2012a) stiller imidlertid spørgsmålstegn herved. I et forløbsstudie fandt forskerne, at skovarbejdere med høj beslutningsmulighed i arbejdet havde en forøget hjerte-kar-dødelighed. Forklaringen på resultatet kan på grund af undersøgelsesdesignet ikke direkte udledes af den gennemførte forskning, men den skal formodentlig netop findes i konteksten (Joensuu et al. 2012b; Rugulies 2012). De finske skovarbejdere blev i 1990'erne udsat for en betydelig omstrukturering af branchen, og det er måske de skovarbejdere med størst beslutningskompetence, som har været hårdest ramt af denne omstrukturering. Dette eksempel illustrerer således problemer med den eksterne validitet, selvom den interne validitet af de mange undersøgelser som påviser den positive sammenhæng mellem indflydelse og helbred, er høj.

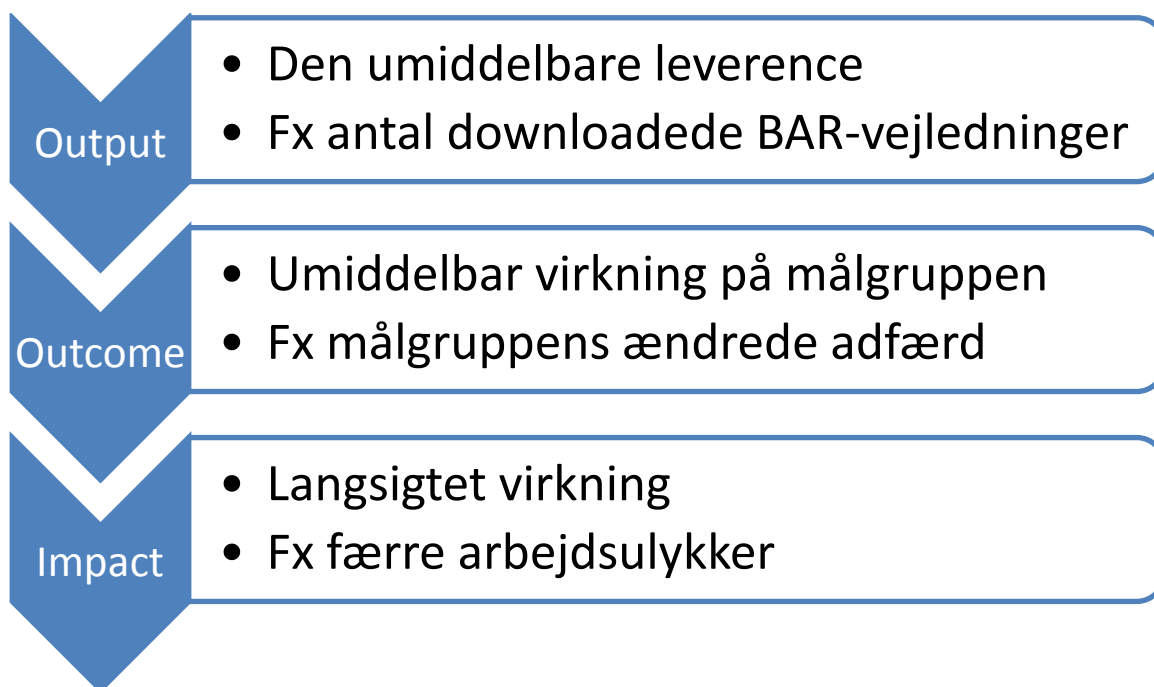
I forlængelse af diskussionen om validitet skal det også nævnes, at der i forskningskredse også opereres med begrebet reliabilitet – eller pålidelighed – når man taler om validitet i forbindelse med en vurdering af forskningskvalitet (Krogstrup, 2011). Det betyder, om andre forskere er i stand til at gentage undersøgelsen og nå frem til de samme resultater. I samfundsmæssige indsatser er det stort set aldrig muligt at gentage den samme indsats, og det er derfor ikke muligt at operere med reliabilitet i traditionel forstand. Det er dog stadig sådan, at kravet om omhyggelig metodebeskrivelse er relevant. Det giver andre forskere, evaluatore og beslutningstagere mulighed for, at vurdere om en undersøgelse metodemæssigt er gennemført på en kvalificeret måde.

### **3.2. Effekt**

I brugen af evidensbegrebet står effekt og årsags-virkningssammenhænge centrale. Vi vil gerne have størst mulig sikkerhed for den ønskede effekt, og vi vil gerne vide mest muligt om, hvordan vi sikrer denne effekt. Men i evalueringssammenhæng er det almindeligt accepteret, at det ofte er meget vanskeligt at sige noget om, hvorvidt slutmålet af en indsats på arbejdsmiljøområdet – i form af mindre nedslidning, færre arbejdsbetingede sygdomme og færre ulykker – er opnået. Sådanne effekter har typisk et meget langsigtet gennemslag og er påvirket af så mange andre faktorer, at det i realiteten er meget vanskeligt at sige noget om denne effekt. Ofte vælges derfor andre evalueringsmål. I bedste fald kan der måles på arbejdsmiljøet. Medarbejdere kan spørges om det psykosociale arbejdsmiljø, eksponeringer kan måles og skærme på

rundsave kan registreres. Det kræver selvfølgelig, at der er velbegrundede teorier om, at de forhold i arbejdsmiljøet man måler, faktisk har betydning for sikkerhed og sundhed. Ofte er det imidlertid også vanskeligt; både fordi fx et BAR-projekt ikke kommer så tæt på den enkelte virksomhed, og fordi det ville blive for dyrt at måle.

Man skelner ofte mellem output, outcome og impact. Output er den umiddelbare 'leverance' af indsatsens kerneaktiviteter, outcome er de virkninger indsatsen har overfor målgruppen på lidt længere sigt, mens impact er den mere langsigtede virkning fx i form af holdbare forandringer. Lad os tage en BAR-vejledning som eksempel. Hvorvidt den er blevet spredt eller downloadet, eller om den er blevet læst kan siges at være output. Kan folk huske hvad de har læst, har de fået noget nyt at vide, er de blevet klogere; og videre har de ændret adfærd; har vejledningen gjort en forskel kan siges at være outcome. Har BAR-vejledningen grundlæggende forbedret arbejdsmiljøet på det lidt længere sigt, har vejledningen påvirket de ansattes sikkerhed og sundhed kan siges at være impact.



**Figur 1: Definitioner på output, outcome og impact af indsatser**

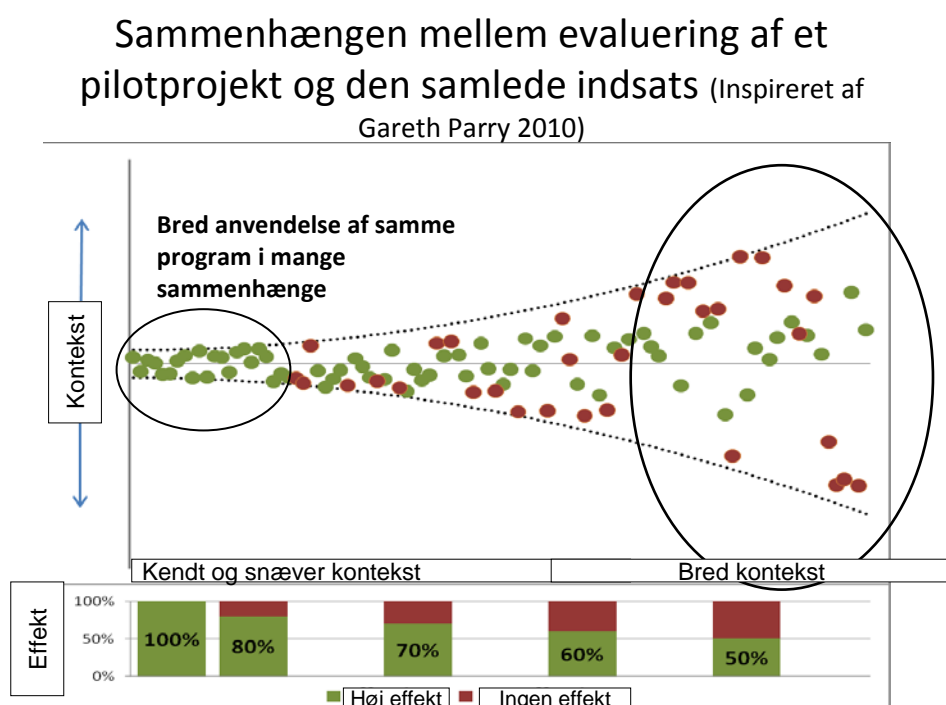
Af pragmatiske årsager bliver det derfor typisk valgt at evaluere på outputs på aktivitetssiden. Det er et udtryk for det umiddelbare produkt, som en given aktivitet har. Det kan fx i forbindelse med informationsindsatser være de umiddelbare tal for antallet af downloads fra en hjemmeside eller antallet af distribuerede pjecer, men det kan også være outputs i næste led fx i form af andel af en gruppe som har kendskab til den information som distribueres gennem hjemmesiden og pjecerne.

Disse outputs må imidlertid ikke forveksles med det ønskede resultat for indsatsen. Der er ikke nødvendigvis sammenhæng mellem kendskab til bestemt viden og handlinger, som forbedrer arbejdsmiljøet, men en god programteori kan måske på basis af sekundær viden fra andre sammenhænge sandsynliggøre, at de målte outputs kan bruges til at vurdere, om den ønskede effekt er nået. Om det kan gøres meningsfyldt, afhænger af den sekundære videns eksterne validitet – altså i hvilken udstrækning den

kan generaliseres fra den kontekst, hvor den er generet, til den anden kontekst, hvor den anvendes i en programteori fx for en bestemt informationsindsats. Dokumentation for evidens som udtryk for at en bestemt indsats eller strategi kan forventes at have den ønskede effekt, er i mange sammenhænge et krav, der stilles inden at indsatsen bringes i bredere anvendelse fx når der udvikles nye medicinske behandlingsformer af sygdomme. Et sådant krav stilles sjældent til arbejdsmiljøindsatser. Erfaringer opsamles oftest efter at en besluttet indsats er afprøvet. I forbindelse med bevillinger til projekter og indsatser som fx i Forebyggelsesfonden forslås det ofte, at der først gennemføres et pilotprojekt. Målet er oftest at kunne justere den store og bredere indsats. Evalueringer kunne med fordel anvendes i forbindelse med sådanne pilotprojekter for på den måde også at teste programteorien og justere de forventede resultater. Herved kan pilotprojekter anvendes til mere end blot at justere indsatsen på de indre linjer, men netop også til at justere indsatsen i forhold til et større kendskab til den kontekst man forventer at udbrede indsatsen i. Herunder fx målgruppens variation, forskellige behov og reaktion på forskellige virkemidler.

I et pilotprojekt vil der ofte være flere ressourcer til de enkelte aktiviteter og en større bevågenhed på målgruppen. Man kan derfor forvente et større gennemslag. Man kan derfor opnå forskellige resultater ved at evaluere et pilotprojekt i forhold til evalueringen af den samlede indsats.

Dette forhold er illustreret i den følgende figur:



**Figur 2 Sammenhæng mellem evaluering af en indsats i en kendt og begrænset kontekst og i en bred kontekst. (Inspireret af Gareth Parry 2010)**

Figuren illustrerer hvordan evaluering af et forholdsvis begrænset antal cases, som fx kunne være omfattet af et pilotprojekt, kan opnå en stor dokumentation for at indsatsen vil have en virkning.



Når den udbredes til en større sammenhæng vil der være både effektive og ikke effektive cases, og vurderingen af indsatsens virkning kan som i figuren falde til en effektivitet på 50 %. Gennemsnittet vil derfor pege på en ringe effekt, men det vigtige er imidlertid, at der er 50 % af målgruppen, som rent faktisk reagerer på samme positive måde som målgruppen i pilotprojektet.

Evaluering af et pilotprojekt fortæller os således noget om, hvordan en indsats virker eller kan virke, hvorimod evalueringen af den samlede indsats fortæller os noget om, overfor hvem og i hvilken kontekst denne virkning kan forventes at indtræffe.

#### **4. Fra evaluering til handling**

Hovedformålet med evaluering er altså at fremskaffe den bedst mulige evidens for fremtidige handlinger. Det er i den sammenhæng problematisk at anvende udtrykket: "Der er evidens for..." Det er vanskeligt – udover banale eksempler – at forestille sig, at der for virkemidler i arbejdsmiljøindsatsen fremskaffes den endegyldige evidens for, hvornår noget virker eller ikke virker. Ovenstående eksempel fra de finske skovarbejdere viser netop, hvordan en ret generaliserbar viden om de positive aspekter af indflydelse også skal vurderes i deres egen sammenhæng.

I realiteten vil beslutningstagere (politikere, myndigheder, organisationer m.v.) derfor altid skulle træffe beslutninger på et ufuldstændigt vidensgrundlag. Et krav om fuld og sikker viden vil kun skabe handlingslammelse. Et eksempel er børneopdragelse: Forældrene står principielt i en vanskelig situation. De vil gerne give deres børn den bedst mulige opdragelse, men selvom de læser alt, hvad der er skrevet om børneopdragelse, vil de aldrig have tilstrækkelig viden til altid at træffe de rigtige beslutninger. Samtidig er de nødt til hver eneste dag at træffe beslutninger. Handlingslammelse vil under alle omstændigheder være næsten den værste tænkelige løsning på problemet.

Viden genereret ved evaluering har imidlertid det indbyggede problem, at den altid er bagudrettet (selvom responsiv evaluering kan have en vis samtidighed). Beslutningen tages, indsatsen sættes i gang, og først når den er afsluttet, vil der foreligge et evalueringresultat. På det tidspunkt har den oprindelige beslutning nærmest kun historisk interesse. Beslutningstagerne står over for nye beslutninger om konkrete indsatser, som i sagens natur endnu ikke er evalueret.

Alt for ofte ender evalueringresultater derfor med ikke at blive anvendt. De tjener derfor blot til at udfylde et sekundært formål med at legitimere den oprindelige beslutning og påvise, at beslutningstagerne optræder forsvarligt ved at gennemføre evaluering af deres indsatser.

Selvom der handles på et ufuldstændigt grundlag, er der stadig både behov for og mulighed for at anvende eksisterende viden. Der vil typisk være en stor mængde viden til rådighed, som både omfatter tidligere evalueringer, forskning og ikke mindst beslutningstagernes praktiske erfaringer både med tidligere indsatser og med at få indsatser tilpasset til noget, som er politisk realiserbart. Som nævnt i notat 2 bygger enhver indsats på en implicit eller eksplicit programteori. Det er netop ved beslutninger på et ufuldstændigt grundlag, at en eksplicit og reflekteret programteori kan spille en vigtig rolle. Gennem arbejdet med at sætte ord på programteorien bringes viden og erfaringer i spil samtidig med at man over for sig selv og andre begrundes, hvorfor man antager, at indsatsen vil virke. Der vil være elementer i programteorien, der er baseret på relativt sikker viden; mens for andre elementer vil den tilgængelige viden være mere usikker

eller manglende. Her må man gennemføre et kvalificeret skøn for at bygge bro mellem de områder, hvor viden er mere sikker (Pawson, Wong, & Owen 2011).

#### **4.1. Interesser bag evaluering**

Dette problem bliver yderligere forstærket af de interesser, som knytter sig til evaluering. Normalt lægger man vægt på at få en uvildig evaluering, men det er i praksis mere et ideal end en realitet. En evaluering bliver typisk sat i gang af de samme beslutningstagere, som er ansvarlige for den indsats, som skal evalueres. Mens dette sammenfald kan give relativt gode muligheder for at anvende evalueringresultater, fordi det er den primære bruger som bestiller evalueringen, er der også en bagside. Beslutningstagerne vil næppe have interesse i evalueringresultater, som påviser, at det var en forkert beslutning, og pengene er spildt. Samtidig er evaluator typisk en konsulentvirksomhed, som lever af at lave evalueringer, og den vil normalt ikke have interesse i at lægge sig ud med bestilleren af evalueringen, samtidig med at evaluator selvfølgelig er afhængig af at bevare sit faglige renommé.

Selvom såvel bestiller som evaluator formodentligt er pålidelige institutioner og enkeltpersoner, som gør deres bedste for, at en evaluering bliver brugbar, så kan interessed spørgsmål ikke undgå at præge evalueringen. Det kan være i større eller mindre grad. I nogle tilfælde er der ikke så mange potentielle problemer knyttet til en evaluering, mens der kan være mere kontroversielle spørgsmål i andre. Men overordnet set må man antage, at evalueringprocessen i et vist omfang påvirkes af de involverede aktørers forståelser og interesser. I beskrivelsen af evalueringsopgaven vil beslutningstageren typisk nedprioritere spørgsmål, som potentielt kan stille de centrale beslutningstagere i dårligt lys fx ved at fokusere på andre aktørers rolle frem for egen rolle. Tilsvarende vil evaluator være tilbøjelig til at nedtone negative resultater, hvis man får indtryk af, at de giver problemer for bestilleren. Det vil ikke ske ved at forfalske resultater, men ved at fremme nogle resultater frem for andre og fx ved at forklare negative resultater med uforudsete kontekstuelle faktorer fx målgruppers økonomiske problemer eller andre aktørers manglende engagement. Det vil sjældent være forhold, som er forkerte, men der sker let en vægtning af materialet som aflaster bestillerens ansvar. Det samme kan gøre sig gældende over for andre betydningsfulde aktører, som har været inddraget i den evaluerede indsats.

Problemet bliver større jo mere omfattende indsatsen er, og dermed jo større evalueringen er, men ved de større evalueringer kan man forsøge at modvirke denne tendens. Det kan fx ske ved at inddrage forskningsinstitutioner, som må antages at være mindre afhængige af bestilleren, selvom forskningsinstitutioner til stadighed bliver stadig mere afhængige af ekstern finansiering og således også kan have en interesse i at bevare gode relationer til bestilleren. Det kan man også forsøge at imødegå ved at inddrage paneler af uafhængige forskere/eksperter, som har begrænsede økonomiske interesser i den pågældende evaluering og i bestilleren. Der kan dog aldrig findes nogen personer eller institutioner, som er helt interessefrie, og samtidig gør den slags ekstra tiltag evalueringen dyrere og mere kompliceret. Det vil derfor typisk være ved mere omfattende evalueringer af store programmer, at sådanne omhyggelige forsøg på at imødegå risikoen for interesse-bias vil være relevant.

#### **4.2. Pragmatisk evaluering**

Beslutningstagere, som skal iværksætte en evaluering, står altså med en vanskelig opgave. Der skal gennemføres evaluering af indsatser over for vilde problemer, og viden til at tilrettelægge en indsats er ufuldstændig. Midlerne til at gennemføre evaluering er begrænsede, og ikke mindst har både

beslutningstager og evaluator større eller mindre interesser i visse udfald af evalueringen. Der skal derfor findes pragmatiske løsninger, hvor 'det bedste ikke skal blive det godes værste fjende', samtidig med at de mulige pragmatiske løsninger ikke ender med at være så meget 'quick and dirty', at resultaterne af evalueringen bliver misvisende og risikerer at gøre mere skade end gavn.

Der vil sjældent entydigt være tale om direkte skadelige evalueringer, men et par tænkte problematiske eksempler kan anvendes som illustration: Hvis man fx måler på outputs, hvilket kan være billigt (fx antal hits på en hjemmeside), kan man måske få et billede af, at en bestemt indsats har vakt interesse og dermed er godt på vej. Imidlertid er de mange klik måske blot udtryk for, at brugere alligevel er inde på en hjemmeside, som de ofte anvender og i den forbindelse blot undersøger om en nyhed om indsatsen er interessant og faktisk konstaterer at det er den ikke. Et andet eksempel kan være et tilfælde, hvor man gennemfører en række pilotforsøg. Her kan en billig evalueringsmetode bestå i at ringe til de projektansvarlige for at undersøge, hvordan det er gået med forsøgene. Men i det tilfælde vil de projektansvarlige typisk have en relativ positiv holdning til forsøgene, netop fordi de selv har et medansvar for at gennemføre forsøget, og man får dermed et positivt billede af fremdriften, som ikke nødvendigvis er retvisende.

Det væsentligste bliver at få tilpasset evalueringens ønsker til de tilgængelige ressourcer. Hvis en indsats er af begrænset omfang, er det måske smartest at tænke i en responsiv selvevaluering, som ikke nødvendigvis dokumenteres med andet end et mødereferat. Her kunne de involverede samles for foretage en fælles vurdering af, hvordan indsatsen forløber, og om der er behov for korrigerende tiltag. Det kan også være, at ressourcerne måske bedre kan anvendes i en grundigere forberedelse af indsatsen, hvor der arbejdes med en eksplicit programteori. På den måde kan eksperter og forskere i nogle tilfælde måske bedre anvendes til rådgivning i denne tidligere fase frem for i en efterfølgende evaluering, som bliver for overfladisk til at have virkelig værdi.

Selv ved større projekter vil ressourcerne normalt også være begrænsede, og der skal derfor ske en prioritering. Der kan typisk være ønsker om 'lidt fra alle hylde'. Lidt responsiv evaluering for at korrigere undervejs, lidt effektevaluering for at få viden om, hvorvidt indsatsen flytter noget og lidt realistisk evaluering for at få indblik i, hvad der virker. I nogle tilfælde kan et sådan setup måske lade sig gøre. En responsiv selvevaluering, nogle simple output-mål og en evaluering af afgrænsede dele af programteorien som det vurderes vil have særlig interesse. Men det er også her, hvor ambitionerne kan skyde over målet, og resultatet bliver for overfladisk til at bidrage med anvendelig viden.

En særlig mulighed, som i dag anvendes i et vist omend beskedent omfang, er forskellige former for tværgående evalueringer. Der er netop ved at blive igangsat en større evaluering af BAR-systemet, som kan tjene som et eksempel herpå, men man kunne også tænke sig sådanne evalueringer anvendt i større omfang – nok i mere fokuserede udgaver. Her kunne man fx i en BAR eller flere BAR'er sammen kigge på sammenlignelige indsatser. Det kunne fx dreje sig om indsatser, som omfatter interaktive hjemmesider eller indsatser, som omfatter møde- og netværksaktiviteter. En sådan tilgang kunne åbne to muligheder. For det første kan der samles flere ressourcer sammen således, at der bliver mulighed for at foretage en mere ambitiøs evaluering. For det andet vil evaluering af flere parallelle indsatser i forskellige kontekster give mere generaliserbar viden om mekanismer og kontekst i denne form for indsatser. I nogle tilfælde kunne det være relevant at iværksætte, det man kan kalde en metaevaluering (Pawson 2006), hvor resultaterne fra et større antal allerede gennemførte evalueringer analyseres og sammenlignes på tværs. Der

produceres således ikke ny empiri, men bygges på allerede eksisterende resultater fra mere projektnære evalueringer. En sådan tilgang til evaluering vil give gode muligheder for at få dybere indsigt i de anvendte virkemidler, deres mekanismer i samspil med kontekst hos modtagerne og ikke mindst deres generaliserbarhed. Et aktuelt eksempel er, at Forebyggelsesfonden har støttet en lang række projekter, der anvender forskellige former for "barfodsaktører" – medarbejdere, der melder sig til i en tidsafgrænset periode at arbejde for at udbrede en særlig forebyggelsesindsats. Det kan fx være forflytningsvejledere, stresscoaches, mentorer eller trivselsambassadører. En tværgående evaluering af denne type projekter kunne give en værdifuld viden om, hvornår denne indsatsstrategi virker, og hvad der kan begrænse dens udbytte.

## 5. Evalueringsscenerier

Som det fremgår af ovenstående vil de praktiske evalueringer altid have en pragmatisk tilgang i forhold til de tilgængelige ressourcer og den politiske virkelighed. Det gælder i den sammenhæng om at holde tungen lige i munden, således at der kan opnås det størst mulige videnskabelige udbytte af en evaluering. Praksis vil imidlertid være meget afhængig af hvilken type indsats, der er tale om. Vi har allerede peget på, at der vil være store forskelle imellem den begrænsede indsats og den omfattende indsats, og i dette afsluttende afsnit åbner vi yderligere for denne diskussion gennem en præsentation af konkrete cases, som kan fungere som illustration af fire overordnet set forskellige typer af arbejdsmiljøindsatser. Målet er at de kan danne grundlag for diskussioner af hvilke former for evaluering der med fordel kan anvendes til de forskellige indsatsstyper.

### 5.1. Den begrænsede indsats

Mange BAR-aktiviteter udgøres af ret begrænsede indsatser. Indsatser til hvilke der bevilges ret små beløb, og hvor de fleste ressourcer anvendes af de organisationer som stiller med repræsentanter til indsatsen. Typiske eksempler er, at der udvikles og iværksættes aktiviteter af informationsmæssig karakter som fx en BAR-vejledning, en informationspjece, et nyt tema på en hjemmeside. Det følgende eksempel handler om udviklingen af en BAR vejledning inden for bygge- og anlægsområdet.

På baggrund af en række afgørelser fra Arbejdstilsynet blev nogle af organisationerne inden for Bygge og Anlæg sammen med BAR Byg's sekretariat opmærksomme, på at der knyttede sig nogle særlige sikkerhedsrisici til opsætning af gitterspær ved tagkonstruktioner på større byggerier, især i forbindelse med opsætning af de første spær.

BAR Byg udarbejdede herefter en vejledning i sikker og forsvarlig montering af Gitterspær, (kaldet "Fakta om opstilling af gitterspær"). Vejledningen fylder fire sider og forklarer med instruktive tegninger, hvordan arbejdet skal udføres. Inden en sådan vejledning udsendes, er den blevet godkendt af Arbejdstilsynet, så den bygger på og inddrager de gældende regler på området. Hvordan kan man evaluere virkningen af en sådan vejledning? Den findes både i papirform og på nettet. Antallet af udleverede eksemplarer, antal downloadede og antallet af rekvirerede eksemplarer kan relativt let dokumenteres. De er således et udtryk for interessen for at søge denne viden, men jo ikke et udtryk for om reglerne følges, eller om vejledningens anbefalinger omsættes til praksis.

Den arbejdsproces, der er tale om, forløber tidsmæssigt sjældent længere end en eller nogle få dage. På et stort byggeri er det således en meget begrænset del af den samlede proces, der er fokus på. Det vil derfor

ikke være praktisk muligt at registrere omfanget af, hvor ofte reglerne følges. Skal brugen af vejledningen således dokumenteres må aktørerne selv inddrages. Det kræver et omfattende registreringssystem, instruktion, påmindelser osv. og vil derfor også i praksis være vanskeligt. Kunne det lade sig gøre ville det imidlertid være en kilde til en vigtig viden om reglerne og anvisningernes brugbarhed i praksis. Det kunne fx også omfatte registrering af nærvæd hændelser knyttet til denne arbejdsfunktion. Et forsøg på en evaluering af vejledningens 'impact' kunne være at registrere om antallet af ulykker, der knytter sig specifikt til arbejdet med at opsætte gitterspær. Det forudsætter, at ulykkesregistreringen er præcis nok til dette, og at der forekommer ulykker nok til, at det vil have en statistisk værdi, hvilket næppe er tilfældet.

Udfordringen her består i at finde frem til meget simple evalueringstilgange, som benytter sig af meget få ressourcer, men alligevel bidrager med relevant viden – det vil sige evidens, som kan anvendes til at forbedre fremtidige indsatser.

## **5.2. Den omfattende indsats**

Den omfattende indsats kan være lettere at have med at gøre, fordi der i sagens natur er flere ressourcer til rådighed. Det gør det alt andet lige lettere at tænke evaluering, men i langt de fleste tilfælde vil ressourcerne stadig være en begrænsende faktor. Normalt anvendes der under 10 % til evaluering, og man skal derfor op på en million kroner for at få blot 100.000 kr. til evaluering, og for dette beløb er der stadig ikke plads til særlig ambitiøse evalueringdesign. Der er derfor behov for klare programteorier som kan danne grundlag for stærkt prioriterede evalueringer.

### **Eksempel på den omfattende indsats**

BAR Social og Sundhed udgav i 2003 et temahæfte med titlen "Arbejdstid og arbejdsmiljø". Temahæftet byggede på den på det pågældende tidspunkt tilgængelige forskningsbaserede viden om sammenhængen mellem arbejdstid og arbejdsmiljø. Problemstillingens karakter lagde ikke op til vejledninger, der kunne pege på, hvordan konkrete problemstillinger ville kunne imødekommes af konkrete og enkle løsninger. Arbejdstiden reguleres gennem en sammenkædning af rammeaftaler, lokale aftaler og reglerne i arbejdsmiljøloven. Hvad der er fornuftigt og muligt at gøre, vil derfor afhænge meget af den konkrete arbejdsplads. Fra BAR Social og Sundheds side besluttede man derfor på grundlag af den indsamlede viden, at skabe en indsats, der tog form af en bred kampagne, som jævnligt kunne "minde målgruppen" om arbejdstidsproblematikken, for derved at inspirere til at problemerne blev prioriteret, analyseret og håndteret på så mange arbejdspladser som muligt inden for BAR'ens område. Ved at temaet blev "highlightet" over for målgruppen med jævne mellemrum og gennem forskellige medier, var det antagelsen, at det ville sætte sig bedre og bedre fast. Tidsmæssigt skulle indsatsen udfoldes over et år, og det var samtidigt ønsket, at de materialer, der blev udviklet, ville være så holdbare, at de også ville finde anvendelse ud over kampagneåret.

I hvert kvartal udsendtes en "pakke" af materialer som fx en pjece, plakater, kort med slogans, tankevækkende udsagn og gode råd, og klistermærker. Hver pakke havde fokus på et særligt tema, som fx balancen mellem arbejde og fritid, sundere natarbejde, den gode vagtplan. Udsendelsen af materialerne blev fulgt op af temamøder i hver region, rettet mod ansatte og ledere i social og sundhedssektoren. Der oprettedes samtidigt en hjemmeside som både præsenterede materialerne og skabte en mulighed for at føre dialog om temaet.

Målet med en indsats af denne type er således at påvirke arbejdspladserne til selv at tage fat på at vurdere sammenhængen mellem arbejdstidens tilrettelæggelse og arbejdsmiljøet. Virkemidlet er generel oplysning - "prædiken" gennem en række forskellige informationsmaterialer, møder og en hjemmeside. Den særlige ide som denne kampagne bygger på, er at pulser indsatser så den rammer målgruppen ad flere omgange, og derfor styrker interessen også selvom budskabet ikke lige når igennem første gang.

I en evaluering af en række af flere BAR'ers indsatser (Nordisk Kommunikation, 2003) lægges der vægt på, at det netop er vigtigt, at indsatserne ikke er korte afgrænsede forløb, da indsatsens effekt derved let går tabt over tid. I forhold til udbredelsen konkluderer evalueringen at distributionen ofte er det svage punkt ved bredere formidlingsindsatser, og foreslår klarere distributionsaftaler mellem Branchearbejdsmiljørådene og medlemsorganisationerne. Videre anbefales en større udnyttelse af internettets muligheder, hvis man ønsker at nå målet med at holde liv i indsatserne over tid og opbygge tættere relationer til målgrupperne.

En direkte formidling af viden, som i eksemplet om sammenhængen mellem arbejdstidens tilrettelæggelse og sundhed, gennem temadage, fyraftensmøder og lignende bliver generelt vurderet positivt af dem som har deltaget i sådanne aktiviteter. Mens dem der ikke har deltaget kun har en meget begrænset mening om indsatsens relevans. Blandt deltagere møder man ikke sjældent ønsket om, at der i højere grad skabes mulighed for dialog med oplægsholdere om konkrete erfaringer.

I forhold til udbredelse af budskaber, som arbejdstidsindsatsen er et eksempel på, peger evalueringen på, at Branchearbejdsmiljørådene i højere grad bør benytte opinionsdannere, fx arbejdsmiljøunderviserne, som formidlingskanaler.

Generelt foreslår denne evaluering af mange omfattende indsatser, at kvaliteten af BAR indsatserne kan øges, hvis der i højere grad end det synes at være tilfældet, benyttes en mere dynamisk strategisk planlægningsmodel med følgende trin: Situationsanalyse, strategiske mål, kritiske succesfaktorer, målepunkter, handlingsplan, evaluering og læring og opfølgning.

### **5.3. Den tværgående evaluering**

Større tværgående evalueringer er ofte sat i gang fra centralt politisk hold. Det gjaldt fx både den tidligere Capacent evaluering af BAR-systemet og den kommende BAR-evaluering. Der er også eksempler på indsatser, der udvikles og gennemføres af en række BAR'er i fællesskab. Et eksempel er kampagnen: "Gravid med job", som havde til formål at forebygge, at gravide forblev i arbejdsfunktioner, der kunne skade dem selv eller fosteret. Kampagnen var iværksat af seks BAR'er (BAR Byg, I-BAR, BAR jord til bord, BAR for service og tjenesteydelser, Grafisk BAR og BAR transport og en gros).

Den byggede på undersøgelser, der havde dokumenteret, at mange aktører, herunder såvel de gravide selv, som læger, jordmødre og arbejdsgivere er usikre på, hvornår det er forsvarligt for en gravid at blive i sit job, samt hvad der kan gøres for at sikre, at arbejdsmiljøet ikke er skadeligt for den gravide og fosteret. Der blev derfor iværksat en stor oplysningskampagne, der skulle kunne nå ud til alle de nævnte målgrupper. Der oprettedes en hjemmeside ([www.Gravidmedjob.dk](http://www.Gravidmedjob.dk)), hvor der er målrettede indgange for hver af målgrupperne. Den gravide kan klikke på "information til dig der er gravid" og her læse om rettigheder, om hvad hun skal være opmærksom på kan skyldes arbejdsmiljøet, og hvad der må anses for normale smerter,

og endelig hvordan arbejdsmiljøet kan tilpasses, så det ikke skader fosteret, osv. Arbejdsgiveren kan via sin indgang læse mere om arbejdsmiljølovens regler og få information om, hvordan man indfører en graviditetspolitik på arbejdspladsen. Læger og jordemødre kan læse alt relevant om love, arbejdsmiljøregler og forhold som en læge/jordemoder skal være opmærksom på i forbindelse med arbejdsmiljø og graviditet. Og som noget helt unikt er der også adgang til en brancheoversigt med branchespecifik information om arbejdsmiljø for gravide. Der findes fx her en generel beskrivelse af arbejdsforholdene for branchen (og de særlige forhold og risici som måtte gøre sig gældende), en checkliste med alle de ting lægen eller jordemoderen bør spørge ind til og få afklaret ved en konsultation. En evaluering af denne kampagne må derfor opdeles i evalueringer rettet mod hver målgruppe for sig, eftersom målgruppernes behov for og motivation til at anvende hjemmesiden afhænger af deres udgangspunkt. Kampagnen giver gravide god og saglig information, ligesom læger og sygeplejersker kan hente viden, som er specifik ift. de problemer gravide muligvis kan støde på, hvis de arbejder inden for en af de brancher, der er med i kampagnen. Man kan derfor tolke omfanget af "hits" fra de forskellige målgrupper på denne hjemmeside som udtryk for interessen og udbredelsen af indsatsen. Det kan også overvejes om, man kan indarbejde en brugerevaluering, der beder alle som anvender hjemmesiden svare på, det kan fx være spørgsmål til, om hjemmesiden indeholder de informationer, de har brug for osv.

Ingen af disse evalueringer giver dog svar på om hjemmesiden lykkes med at blive et aktivt og brugbart redskab for de konkrete målgrupper. Kvalitative interview viste fx at læger, der havde forsøgt at anvende siden på baggrund af en sag med en gravid, der ikke arbejdede inden for en af de seks medvirkende brancher, tabte interessen for at bruge hjemmesiden, da hjemmesiden mistede legitimitet ved ikke at være bredt dækkende.

#### **5.4. Et program**

Et program er en stor samlet indsats – oftest baseret på en politisk fastlagt prioritering og en i denne sammenhæng afsat økonomisk ramme. De specifikke aktiviteter som udgør programmet er oftest en række forskellige aktiviteter, som er koordineret og løber over en længere periode. Et konkret eksempel er 2020-planen, hvor flere af de enkelte aktiviteter i planen næsten kan beskrives som et program i sig selv, det gælder fx indsatsen rettet mod børn og unge. Et andet eksempel er Forebyggelsesfonden, hvor man kan sige, at programmet er beslutningen om at yde økonomisk støtte til virksomheder, der udvikler konkrete indsatser, som kan forebygge nedslidning eller fremme sundhed. Som en del af programmet indgår hele opbygningen af Forebyggelsesfonden og de økonomiske midler, der er afsat til at drive fonden.

Evaluering af programmer er i sagens natur en kompliceret opgave, der må omfatte mange forskellige evalueringstilgange og metoder. I det følgende giver vi et eksempel herpå, ved at se på, hvordan man fra Forebyggelsesfondens side selv har opstillet rammerne for en evaluering.

I 2011 formulerede Forebyggelsesfonden i et udbud, hvad man forestillede sig at en evaluering af Forebyggelsesfondens virksomhed og betydning skulle omfatte. Man ønskede en evaluering af samtlige af de projekter inden for Fondens fire hovedformål, som var blevet støttet siden oprettelsen af Forebyggelsesfonden og som man forventede ville blive afsluttet senest i 2015. Desuden skulle en ny tilskudsform: Forebyggelsespakkerne evalueres. I udbuddet var der redegjort for, at der ønskedes brug af både proces- og effektevaluering i løsning af evalueringsopgaverne, hvis andre evalueringsformer skulle inddrages skulle evaluator argumentere for dette. I forhold til projekterne inden for de fire hovedformål

ønskede man i udbuddet, at dem der skulle stå for evalueringen også gjorde rede for, om de ville kunne gennemføre en egentlig effektevaluering med kontrolgrupper.

Man ønskede en lærende evaluering, tilrettelagt sådan at evalueringsindsatsen kunne forbedres undervejs for at opnå den mest optimale evalueringsmetode.

Formålet med evalueringen kunne tolkes således, at hovedformålet var, at Forebyggelsesfonden ønskede at lære, hvad der virker godt, og hvad der ikke virker i de gennemførte projekter, herunder hvilke virkemidler, metoder, koncepter og tilgange, projekterne har valgt at anvende, hvilke rammebetingelser projekterne har haft og hvilke effekter, der kan måles i de enkelte projekter. Evalueringen forventede man på den ene side at kunne bidrage til at sprede gode eksempler til andre arbejdspladser og jobgrupper, på den anden side – og måske vigtigst – skulle den kvalificere Forebyggelsesfondens arbejde i forhold til fremtidige udbud, indsatsområder og ikke mindst kunne bidrage til at udvikle større og mere præcise krav til projekternes egen evaluering. Endelig er der også tillagt evalueringens udbud et mål om, at evalueringen parallelt med de andre formål kan klargøre, om man kan identificere samfundsmæssige betydninger af Forebyggelsesfondens indsats. På mange måder er dette den egentlige programevaluering.

Hovedvægten i evalueringen ligger således på evalueringen af de talrige enkelte projekter og de mulige tværgående analyser. Selve programevalueringen - den overordnede evaluering af om Forebyggelsesfonden gennem de støttede projekter og forebyggelsespakker har formålet at have en samfundsmæssig positiv virkning på arbejdsmiljøet - må ifølge udbuddet antages at få en mindre placering i evalueringsopgaven.

Evalueringen opbygges således at evaluator gennem evalueringen af de mange enkelte projekter, efterfølgende bliver i stand til at kunne gennemføre en række analyser på tværs af de mange evaluerede projekter, og herved nå frem til generaliserede konklusioner, om hvad der virker og ikke virker i projekterne.

Evalueringsopgaven bygger på en projektlogisk tankegang, hvor man antager, at man ved at se på projekternes interne sammenhæng mellem den gennemførte indsats og de opnåede effekter, kan vurdere om de er lykkedes med at opfylde deres hensigter. Det kan afdækkes gennem procesevaluering med fokus på det enkelte projekts gennemførelse (ressourcer, aktiviteter og præstationer) og med effektevaluering med fokus på projekternes kortsigtede effekter, der i de enkelte projekter er beskrevet som succeskriterier og projektmål. Disse omfatter fx nedbringelse af sygefravær, reduceret personaleomsætning og øget trivsel.

For at kunne dette forudsætter det, at evaluator kan identificere en programteori for projekterne. Det bliver en opgave for evaluator, at sikre sig at der kan formuleres en programteori. Projekterne analyseres derfor ud fra et spørgsmål til hvert enkelt projekt, om hvorfor projektets valgte indsats skulle kunne føre til de ønskede ændringer for målgruppen?

Skulle der inddrages en kontrolgruppe kunne den indgå ved at projekternes før og eftermålinger sammenlignes med en bred repræsentativ undersøgelse som f.eks. NAK-databasen. En anden form for kontrol ville være mulig gennem kvalitative interview med aktørerne i de enkelte projekter omkring projekternes gennemførelse og effekt. Her må omfanget dog nødvendigvis være begrænset af de tilgængelige ressourcer.



Evalueringensudbuddet lægger som tidligere nævnt vægt på ønsket om, at der gennemføres en lærende evaluering, så erfaringer fra evalueringsopgaverne løbende bruges til at udvikle evalueringensdesignet.

Evalueringsopgaverne vil således i praksis skulle udføres som en kombination af desktopstudier baseret på registrering af de mange projekters projektansøgninger og egen-evalueringer, telefoninterviews samt besøg og interview på et realistisk antal projekter.

Procesevaluering baseret på projekternes programteori og de opnåede resultater kan man også forestille sig kan udføres vha. interview undervejs og ved projekternes afslutning. Det kan bidrage til en analyse af særlige projekttyper, særlige indsatsformer og metoder m.m.

Den egentlige effekt af de mange projekter og dermed af Forebyggelsesfonden som sådan, forestiller man sig i udbuddet skal måles gennem før- og eftermålinger foretaget blandt målgruppen i de støttede projekter. Før- og eftermålinger kan tilrettelægges, så de anvender samme spørgsmål som i store tværsnitsundersøgelser som fx NAK (den Nationale Arbejdsmiljøkohorte) der så kan fungere som kontrolgruppe for før- og eftermålingen. Det vil i denne sammenhæng være væsentligt at overveje om en sådan evaluering bør basere sig på før- og eftermålinger ud fra spørgeskemaer til virksomheder og medarbejdere, eller også skal inddrage spørgeskemaer til facilitatorer og virksomhedsledere kombineret med interviews på udvalgte virksomheder ved projektets afslutning og fx igen et halvt år efter projekternes afslutning.

Det spørgsmål som står tilbage er, hvordan en evaluering af en så stor mængde meget forskellige projekter kan bidrage til en større forståelse af, hvilken samlet virkning på arbejdsmiljøet Forebyggelsesfonden har. Et program som er kendetegnet ved et stort økonomisk omfang, ved at basere sig på virkemidlet økonomisk støtte, og ved at støtte meget forskellige typer af indsatser.

## 6. Konklusion

Dette notat beskæftiger sig med, hvordan evalueringer kan skabe viden om en given indsats virker, og hvis den virker, hvordan den virker. I det første notat skrev vi, at evaluering er en systematisk vurdering af aktiviteter med henblik på at få grundlag for at træffe fremtidige beslutninger. Vores ønsker om evalueringer er således grundlæggende set et ønske om (mere systematisk) at lære, hvad vi skal gøre og hvad vi ikke skal gøre, når vi gerne vil forbedre arbejdsmiljøet.

Vi har gennem de tre notater beskæftiget os med de forskellige problemer, der kan være knyttet til evalueringer. Disse problemer knytter sig til det forhold, at arbejdsmiljøproblemer er 'vilde'. Det skal forstås på den måde, at arbejdsmiljøproblemer udfoldes i komplekse kontekster, der gør det vanskeligt – eller i hvert fald indebærer et omfattende og ressourcekrævende arbejde - at give fyldestgørende svar på sammenhængen mellem årsager og virkninger. I dette notat beskæftiger vi os med de forskellige problemer, der er knyttet til at foretage (gode) evalueringer, og hvordan man så, på trods af problemerne, kan designe kvalificerede evalueringforløb.

De problemer, der knytter sig til evalueringer, hænger sammen med den *viden* evalueringer overhovedet *kan* skabe, den måde man får formuleret det, der skal evalueres (genstanden), de antagelser man evaluerer ud fra og de praktiske forhold man som evaluator eller igangsætter af en evaluering er underlagt.

Evalueringer er generelt bagudrettede. Det vil sige, at de baserer sig på de omstændigheder og den viden, der var til stede ved indsatsens igangsættelse. Og da nogle indsats kan strække sig over en årrække, kan det betyde, at en evaluering, der selvom den er omfattende, grundig og ressourcetung, kan opfattes som irrelevant. Evalueringer producerer på den måde nogen gange viden, der kan opfattes som irrelevant, alene fordi der er et tidsmæssigt slæk mellem start og slut, og den politiske dagsorden i mellemtiden har ændret sig.

Evalueringer undersøger som udgangspunkt om noget virker - altså om det har en effekt. Men en effekt kan man iagttage på mange forskellige måder. Vi har i notatet set på forskellene mellem at se effekt som output, outcome og impact. Output er den umiddelbare 'leverance' af indsatsens kerneaktiviteter. Outcome er de virkninger indsatsen har overfor målgruppen på lidt længere sigt, mens impact er den mere langsigtede virkning fx i form af holdbare forandringer.

Evaluerings kvaliteten er, som vi også behandlede omfattende i det andet notat, afhængige af de antagelser, der lægges til grund for udførelsen af evalueringen. I det andet notat diskuterede vi programteorier som det samlede sæt af antagelser om, hvordan en indsats vil virke. I dette notat har vi i højere grad beskæftiget os med, de antagelser der generelt set ligger til grund for design og anvendelse af evalueringer.

Evalueringer og deres popularitet skal ses i sammenhæng med den store opmærksomhed, der er på evidens. Evidens er et begreb, som bruges og misbruges til mange forskellige formål. I sin grundbetydning betyder det blot, at vi kan underbygge vores antagelser med oplysninger. Inden for forskningen trækker evidensbegrebet dog en noget mere vægtig og ophøjet betydning – som viden, der er videnskabeligt underbygget. I den forbindelse bruges evidensbegrebet også i særdeleshed inden for det medicinske område, hvor vi kender det fra både Cochrane- og Campbell-tilgangene. Begge tilgange favoriserer de såkaldte Randomiserede Control Trials (RCTs), hvor det er tilfældigt (randomiseret), hvem der modtager behandlingen eller indsatsen, og hvor effekten måles både hos modtagerne af behandlingen/indsatsen og hos en kontrolgruppe, som ikke modtog den. Man bruger altså i den forbindelse evidens som begreb i kontekster, der i arbejdsmiljøsammenhænge er mere 'tamme' end 'vilde'. Konteksten i arbejdsmiljøsammenhænge har stor betydning, jf. eksemplet med rundsaven, hvor forebyggelse ikke kun afhænger af afskærmningen, men også af arbejdspladsens sikkerhedskultur. Inden for den medicinske måde at bruge evidensbegrebet forsøger man i højere grad at afkoble netop de variable, der i en arbejdsmiljømæssig sammenhæng viser sig at være de interessante og besværlige.

Når man inden for arbejdsmiljøforskningen og evalueringen af 'vilde' problemer arbejder med evidens, skal det altså foregå på lidt andre måder, end man kan gøre det i den medicinske forskning. Og dette betyder også at nogle af vores gængse forestillinger om undersøgelsers troværdighed (reliabilitet) og validitet sættes under pres. Man deler traditionelt validiteten op i intern og ekstern validitet. Den interne validitet handler om den anvendte metode (teoretisk og praktisk) faktisk er i stand til at påvise de sammenhænge, som man vil undersøge. Mens den eksterne validitet handler om mulighederne for at generalisere resultatet til en bredere sammenhæng. I samfundsmæssige indsats er det stort set aldrig muligt at gentage den samme indsats, og det er derfor ikke muligt at operere med reliabilitet i traditionel forstand. Det er dog stadig sådan at kravet om omhyggelig metodebeskrivelse er relevant. Det giver andre forskere, evaluatore og beslutningstagere mulighed for at vurdere om en undersøgelse metodemæssigt er gennemført på en kvalificeret måde.

Endelig er valget af evalueringsdesign som regel begrundet i de tilgængelige ressourcer og den politiske virkelighed evalueringen skal foretages i. Ofte er det antagelsen at jo mere evalueringerne lever op til de idealer, man finder inden for den naturvidenskabelige forskning, des bedre. Det er imidlertid sådan, at fx uvildige evalueringer mere er et ideal end en realitet i praksis. Det skyldes, at evalueringer typisk sættes i gang af de samme beslutningstagere, som er ansvarlige for den indsats, som skal evalueres. Mens dette sammenfald kan give relativt gode muligheder for at anvende evalueringresultater, fordi det er den primære bruger som bestiller evalueringen, er der også en bagside. Beslutningstagerne vil næppe have interesse i evalueringresultater, som påviser, at det var en forkert beslutning, og pengene er spildt. Samtidig er evaluator typisk en konsulentvirksomhed, som lever af at lave evalueringer, og den vil normalt ikke have interesse i lægge sig ud med bestilleren af evalueringen, samtidig med at evaluator selvfølgelig er afhængig af at bevare sit faglige renommé.

Mange evalueringer ender derfor med at være såkaldt pragmatiske evalueringer, der plukker lidt hist og her blandt de forskellige evalueringstilgange. Det kan der være gode begrundelser for, men der er også en oplagt risiko for at sætte sig mellem flere stole. Den information, som kommer ud af en sammensat evaluering med ønsker om lidt fra alle hylder, kan være så usikker at den reelt ikke bidrager til et forbedret beslutningsgrundlag.

Udover at valget af evalueringsdesign er begrundet med det praktisk mulige, er det også meget afhængig af hvilken type arbejdsmiljøindsats, der er tale om. Den sidste del af notatet handler netop om, hvordan man kan sammensætte (pragmatiske) evalueringssindsatser, alt efter hvilken type indsats, der er tale om. Vi arbejder med fire forskellige evalueringsscenarier som hver for sig på forskellig måde skal tilpasses til de tilgængelige ressourcer.

#### 1. Den begrænsede indsats

Evalueringssdesignet skal her være simpelt og benytte meget få ressourcer, der alligevel bidrager med relevant viden.

#### 2. Den omfattende indsats

Evalueringssdesignet må her bygge på klare programteorier som kan danne grundlag for stærkt prioriterede evalueringer. Der bør udformes en dynamisk strategisk planlægningsmodel med følgende trin: Situationsanalyse, strategiske mål, kritiske succesfaktorer, målepunkter, handlingsplan, evaluering og læring og opfølgning.

#### 3. Tværgående evaluering

Evalueringssdesignet må her forholde sig til, at indsatsen sker i flere forskellige målgrupper, hvorfor designet udover at forholde sig til flere komplekse programteorier også må være særligt opmærksom på de forskellige målgrupper

#### 4. Et program

Et program består typisk af flere projekter, hvorfor evalueringen må opbygges således at evaluator gennem evalueringen af de mange enkelte projekter, bliver i stand til at kunne gennemføre tværgående analyser af de evaluerede projekter, og herved nå frem til generaliserede konklusioner, om hvad der virker og ikke virker i projekterne. Også her må evaluator være opmærksom på en reflekteret programteori for de enkelte projekter. Projekterne analyseres derfor ud fra et spørgsmål til hvert enkelt projekt om, hvorfor projektets valgte indsats skulle kunne føre til de ønskede ændringer for målgruppen

## Litteratur

Dahler-Larsen, P. 2003, Selvevalueringens hvide segl Syddansk Universitetsforlag, Odense.

Davies, T. O., Nutley, S. M. and Smith, P. C. (2000). What Works: Evidence-based Policy and Practice in the Social Services. Bristol: The Policy Press.

Hargreaves, David H.: In Defence of Research for Evidence-based Teaching: a rejoinder to Martyn Hammersley, British Educational Research Journal, Volume 23, Issue 4, September 1997, pages 405-419

Joensuu, M., Kivimäki, M., Koskinen, A., Kouvonen, A., Pulkki-Råback, L., Vahtera, J., Virtanen, M., & Väänänen, A. 2012a, "Differential Associations of Job Control Components With Mortality: A Cohort Study, 1986-2005", American Journal of Epidemiology, vol. 175, no. 7, pp. 609-619.

Joensuu, M., Kivimäki, M., Koskinen, A., Kouvonen, A., Pulkki-Råback, L., Vahtera, J., Virtanen, M., & Väänänen, A. 2012b, "Joensuu et al. Respond to "Structure and Context Matters"", American Journal of Epidemiology, vol. 175, no. 7, pp. 625-626.

Krogstrup, H. K. 2011, Kampen om evidens - resultatmåling, effektevaluering og evidens Hans Reitzels Forlag, København.

Nordisk Kommunikation: Evaluering af fire indsatser gennemført af Branchearbejdsmiljørådene Social & Sundhed, Undervisning & Forskning samt Finans/Offentlig Kontor & Administration. Hovedrapport 2003. Udgivet af Arbejdsmiljøsekretariatet.

Parry, Gareth 2010: Building and Assessing Evidence through the Science of Improvement. Power point presentation from Institute for Healthcare Improvement, Harvard medical School. Pawson, R. 2006, Evidence-based policy: A realist perspective Sage.

Pawson, R., Wong, G., & Owen, L. 2011, "Known Knowns, Known Unknowns, Unknown Unknowns: The Predicament of Evidence-Based Policy", American Journal of Evaluation, vol. 32, no. 4, pp. 518-546.

Rieper, Olaf og Hanne Foss Hansen (2007): Metoddebatten om evidens. AKF Forlaget.

Rugulies, R. 2012, "Invited Commentary: Structure and Context Matters: The Need to Emphasize "Social" in "Psychosocial Epidemiology"", American Journal of Epidemiology, vol. 175, no. 7, pp. 620-624.

Simons, Helen: Evidence-based practice: panacea or over promise? Research Papers in Education; Dec. 2003, Vol. 18 Issue 4, p303-311.