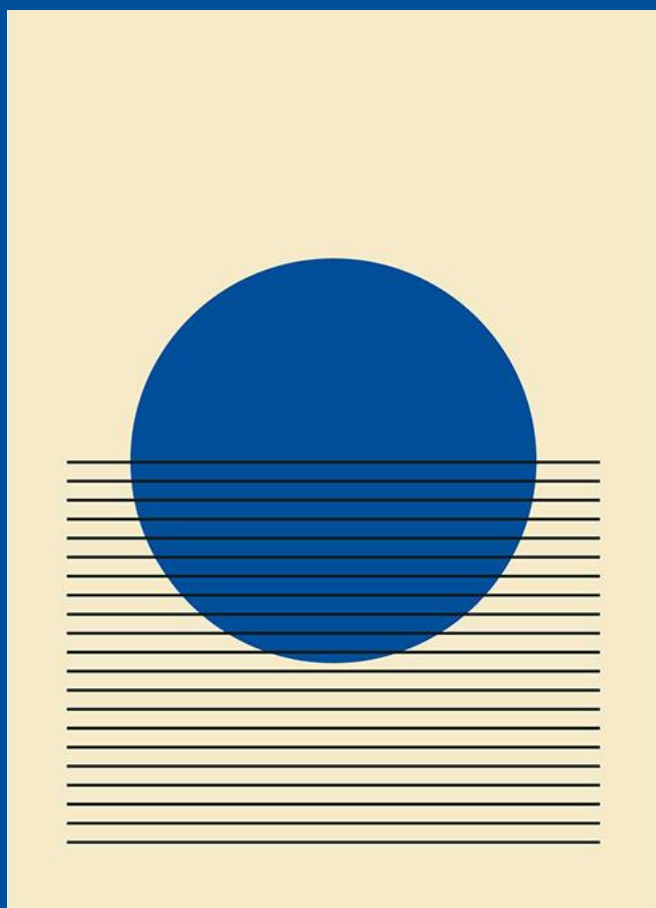


Ny indsats og funktionalitet til prototypen af regnemodel for uddannelsesinvesteringer (REFUD)

Tillæg til teknisk dokumentationsrapport



*Ny indsats og funktionalitet til prototypen af regnemodel for
uddannelsesinvesteringer (REFUD)
– Tillæg til teknisk dokumentationsrapport*

© VIVE og forfatterne, 2022

e-ISBN: 978-87-7582-042-9

Modelfoto: Ricky John Molloy/VIVE

Projekt: 302180

Finansiering: Børne- og Undervisningsministeriet

VIVE

Det Nationale Forsknings- og Analysecenter for Velfærd

Herluf Trolles Gade 11

1052 København K

www.vive.dk

VIVEs publikationer kan frit citeres med tydelig kildeangivelse.



Forord

Denne rapport er et tillæg til den tekniske dokumentationsrapport udarbejdet i forbindelse med udviklingen af en generisk prototype af regnemodel for uddannelsesinvesteringer (REFUD). Arbejdet udspringer af et indledende arbejde, hvor VIVE afdækkede mulighederne for udbredelse af SØM til indsatser på børne- og undervisningsområdet samt arbejdet med udviklingen af en prototype. Arbejdet og rapporten er finansieret af Børne- og Undervisningsministeriet.

REFUD-prototypen består af et beregningsmodul og inputdata udarbejdet via programmer på Danmarks Statistiks forskermaskine, mens empirien vedrørende indsats-effekter er hentet fra et studie i forskningslitteraturen.

Rapporten er skrevet af senioranalytiker Karl Fritjof Krasel, seniorforsker Rune Vammen Lesner, forsker Anne Brink Nandrup samt videnskabelig assistent Matvei Andersen.

Rapporten har været læst og kommenteret af ekstern reviewer, ligesom den har gennemgået VIVEs interne kvalitetssikring. Reviewere takkes for gode og brugbare kommentarer.

Hans Hummelgaard

Forsknings- og analysechef for VIVE Effektmåling



Indholdsfortegnelse

Sammenfatning	5
---------------	---

1	Indledning	7
---	------------	---

2	Vægtningsmodul	8
2.1	Klassificering af studier	8
2.2	Fra klassificering til vægte	11
2.3	Vægtede estimater	12
2.4	Evidensniveau	13
2.5	Robusthedsanalyse	14

3	Multiple typer af evidens og mekanisme til prioritering	22
3.1	Direkte evidens	22
3.2	Multipel evidens og prioritering	23

4	Ny indsats: Klassekvotient	25
4.1	Helårseffekter	25
4.2	Målgruppe og depreciering	25

5	Nyt link outcome: Elevtrivsel	27
5.1	Konstruktion af link outcomes	28
5.2	Datatilgængelighed	29

Litteratur	31
------------	----

Bilag 1 Ordforklaringer	32
-------------------------	----

Bilag 2 Vægte	33
---------------	----

Bilag 3 Oversigt over klassificering af studier	34
---	----

Sammenfatning

Formålet med dette projekt er at videreudvikle en prototype af regnemodel for uddannelsesinvesteringer (REFUD). Arbejdet har taget udgangspunkt i 'Udvikling af regnemodel for uddannelsesinvesteringer (REFUD) - Teknisk dokumentationsrapport' (Lesner, Krassel, & Nandrup, 2022). Det anbefales derfor, at denne tillægsrapport læses i forlængelse af 'Udvikling af regnemodel for uddannelsesinvesteringer (REFUD) - Teknisk dokumentationsrapport'.

Projektet indeholder i alt fem delleverancer:

1. Dette tillæg til den tekniske dokumentationsrapport
2. Et opdateret beregningsmodul
3. Opdaterede programmer til bearbejdning af inputdata på Danmarks Statistiks forskermaskine og modelinput
4. Et vægtningsmodul, der gør det muligt at sammenfatte viden fra forskningslitteraturen
5. Protokol for indskrivning i REFUD-vægningsdatabase.

Rapportens indhold

Dette tillæg til den tekniske dokumentationsrapport beskriver udvidelserne af modellen med hensyn til:

- > vægtningsmodul
- > multiple typer af evidens og mekanisme til prioritering
- > ny indsats: klassekvotient
- > nyt link outcome: elevtrivsel.

Hver af disse beskrives nedenfor.

Vægningsmodul

Effektstørrelsen, der angiver effekten af indsatsen på link outcome, er en meget vigtig del af den samlede beregning i regnemodel for uddannelsesinvesteringer (REFUD). Effektstørrelserne til beregningen kommer fra eksterne kilder i form af effektmålinger, der har været gennemført med forskellige designs, forskellige populationer og i forskellige kontekster. Endvidere vil der i nogle tilfælde være få eller måske slet ingen effektstørrelser til rådighed, mens der i andre tilfælde vil være mange – alt afhængig af størrelsen af den akademiske

litteratur på det specifikke område. Vægtningsmodulet gør det muligt at inddrage viden fra en større del af forskningslitteraturen i samme beregning. Dette gøres ved at vægte studierne efter deres relevans for modellen. I tillæg tilbyder modulet et mål for det tilgængelige evidensniveau for en given indsats.

Multiple typer af evidens og mekanisme til prioritering

Beregningsmodulet til modellen udvides til at kunne håndtere multiple typer af evidens. Det vil sige, at det opdaterede beregningsmodul kan håndtere multiple link outcomes og direkte evidens inden for samme indsats. Flere evidenskilder medfører et behov for, at beregningsmodulet skal være i stand til at prioritere mellem disse. Beregningsmodulet udvikles derfor således med en prioriteringsmekanisme, der gør det muligt at vælge forskellige prioriteringsmodeller.

Ny indsats: klassekvotient

Rapporten beskriver, hvorledes indsatsen Klassekvotient konkretiseres, således at den kan implementeres i modellen. I modellen defineres klassekvotientindsatsen som effekten af at reducere klassestørrelsen med én elev i alle årets undervisningstimer. Modellen giver dog mulighed for at regne på mere eller mindre intensive indsatser.

Nyt link outcome: elevtrivsel

Elevtrivsel i folkeskolen tilføjes som link outcome i modellen ved brug af Den Nationale Trivselsmåling for 4.-9. klasserne. Specifikt tilføjes fem link outcomes:

1. Trivselsmål for faglig trivsel
2. Trivselsmål for social trivsel
3. Trivselsmål for støtte og inspiration
4. Trivselsmål for ro og orden
5. Et samlet trivselsmål for generel trivsel.

De fem mål har til formål at indfange forskellige aspekter af elevtrivsel. Definitionen af målene følger den almindelige tilgang i BUVM (se (Styrelsen for It og Læring, 2017) for en detaljeret beskrivelse).

1 Indledning

Regnemodel for uddannelsesinvesteringer (REFUD) har til formål at gøre brugeren i stand til at beregne socialøkonomiske konsekvenser med et fireårigt perspektiv af en given indsats inden for børne- og undervisningsområdet.

I rapporten 'Afdækning af muligheder for udbredelse af SØM til indsatser på børne- og undervisningsområdet' (Jacobsen et al., 2022) afdækkes VIVE mulighederne for at udbrede principperne bag SØM til indsatsområder på børne- og undervisningsområdet og i 'Udvikling af regnemodel for uddannelsesinvesteringer (REFUD) - Teknisk dokumentationsrapport' (Lesner et al., 2022) beskrives, hvorledes dette gøres i praksis. Denne rapport skal læses som et tillæg til (Lesner et al., 2022).

Metodisk står modellen på to ben. Det ene ben er det empiriske datafundament og det andet ben er evidens fra forskningslitteraturen. Formålet med dette projekt er at videreudvikle modellen både med hensyn til datafundament og benyttelsen af evidens fra forskningslitteraturen samt at videreudvikle selve beregningsmodulet.

Datafundamentet udvides ved at inddrage information om børn og unges trivsel i skolen og fra Den Nationale Trivselsmåling i 4.-9. klasse. Dette gøres efter ønske fra BUVM i en forståelse af, at ændring i elevtrivsel kan være et væsentligt udfald af indsatser på børne- og undervisningsområdet. Derved kan elevtrivsel benyttes som link outcome i modellen.¹

Benyttelsen af evidens fra forskningslitteraturen udvikles ved at introducere et vægtningsmodul, der gør det muligt at inddrage viden fra en større mængde af forskningslitteraturen i modellen.

Selve beregningsmodulet udvides med en ny indsats, Klassekvotient, samt med en ny funktionalitet, der gør det muligt at inddrage evidens fra flere kilder samtidig samt prioritere mellem disse.

Rapporten er organiseret som følger: I kapitel 2 beskrives vægtningsmodulet. Dernæst følger i kapitel 3 en beskrivelse af, hvorledes modellen kan benytte multipel evidens og prioriterer imellem disse, mens kapitel 4 beskriver implementeringen af indsatsen Klassekvotient i modellen. Kapitel 5 beskriver implementeringen af elevtrivsel som link outcome.

Som en hjælp til læseren findes i **Fejl! Henvissningskilde ikke fundet.** ordforklaringer af rapportens ord og begreber.

¹ Se Lesner et al. (2022) for en beskrivelse af modellen og herunder link outcomes.

2 Vægtningsmodul

Effektstørrelsen, der angiver effekten af indsatsen på link outcome, er en meget vigtig del af den samlede beregning i REFUD. Effektstørrelserne til beregningen kommer fra eksterne kilder i form af effektmålinger, der har været gennemført med forskellig designs, forskellige populationer og i forskellige kontekster. Endvidere vil der i nogle tilfælde være få eller måske slet ingen effektstørrelser til rådighed, mens der i andre tilfælde vil være mange – alt afhængig af størrelsen af den akademiske litteratur på det specifikke område. Vægtningsmodulet tilbyder et mål for det tilgængelige evidensniveau for en given indsats.

Nedenfor angives, hvilke principper og metoder modulet anvender ved udvælgelse af den effektstørrelse, modellen skal regne med. Dette beskrives nedenfor.

2.1 Klassificering af studier

I REFUD klassificeres studier baseret på en af VIVE udfærdiget protokol (se Lesner & Nandrup, 2022). Protokollen beskriver, hvordan hvert studie skal registreres i en database, således at en kvalitetsvurdering af det enkelte studie og en vægtning på tværs af studier gøres muligt. Databasen vil derved indeholde et flerdimensionelt mål for det enkelte studies relevans for modellen.

Hvert studie klassificeres på baggrund af følgende *hovedfaktorer*:

6. Kvalitet
7. Kontekst

Sammenlignelighed.

Tabel 2.1 Hovedfaktorer for klassificering af studier til brug i REFUD

Kvalitet	Kontekst	Sammenlignelighed
Udgivelsesniveau	Geografisk område	Indsats
Empirisk metode	Tidspunkt for indsats	Outcome
Datagrundlag	Populations repræsentativitet	Tidshorisont

Hovedfaktorerne uddybes i det følgende.

Den forskningsmæssige kvalitet af et studie kan være mangfoldig. Her defineres studiets kvalitet alene på baggrund af dets evne til at identificere en kausal sammenhæng mellem en modelrelevant indsats og et eller flere modelrelevante outcomes. Dette gøres ved brug af følgende tre *subfaktorer*:

1.1 Udgivelsesniveau

1.2 Empirisk metode

1.3 Datagrundlag.

Udgivelsesniveauet af et studie baseres på niveauet af det videnskabelige tidsskrift, analysen er udgivet i. Niveauet måles konkret ved at opdele tidsskrifter i niveauer baseret på deres BFI, hvor det højeste niveau er gruppen med det højeste BFI-niveau²³. Udgivelsesniveauet vil således være et mål for den forskningsfaglige vurdering af studiets kvalitet. Denne vurdering kan dog ikke stå alene, da forskningsfaglige vurderinger baseret på udgivelsesniveau blandt andet også vil vægte nye teoretiske landevindinger og 'first mover'-studier højt. Vi er derfor nødsaget til at tilføje kvalitetsmål på analysens metode og datagrundlag. Den empiriske metode vurderes ved fire niveauer baseret på metodens evne til at estimere en kausal sammenhæng, alt andet lige. Rangeringen foretages på baggrund af et klassisk kvantitativt evidenshierarki som følger: 1) lodtrækningsforsøg, 2) kvasi-eksperimentelle studier, 3) panelmodeller og 4) andet. Med studiets datagrundlag vurderes, hvorvidt studiet har et tilstrækkeligt datagrundlag eller ej. Formålet med denne vurdering er alene at sortere meget små studier fra, da det skønnes, at ekstern validitet i disse studier er begrænset.

I forhold til anvendeligheden af beregningerne foretaget med REFUD er det imidlertid ikke tilstrækkeligt alene at vurdere studiets kvalitet. For at opnå et brugbart mål for studiets modelrelevans bør *studiets kontekst* også tages i betragtning. VIVE foreslår at studiets kontekst vurderes ud fra:

2.1 Geografisk område

2.2 Tidspunkt for indsats

2.3 Populations-repræsentativitet.

² Den Bibliometriske Forskningsindikator (BFI) opdateres løbende og er tilgængelig på Uddannelses- og Forskningsministeriets hjemmeside (<https://ufm.dk/forskning-og-innovation/statistik-og-analyser/den-bibliometriske-forskningsindikator/BFI-lister>).

³ I tilfælde, hvor en analyse i en offentlig tilgængelig rapport vurderes til at være af særlig høj kvalitet, men hvor resultaterne ikke er udgivet i et videnskabeligt tidsskrift, da tildeles studiet et BFI-niveau i vægtningen på baggrund af en faglig vurdering af sammenlignelige studier i videnskabeligt tidsskrifter.

Det optimale studie er således et studie udført i en nyere dansk kontekst, betinget på studiets kvalitet. Geografiske områder rangeres på baggrund af, i hvor høj grad skole-daginstitutionsområdet er sammenligneligt med en dansk kontekst. Vi vil således skelne mellem danske, nordiske og andre studier. Tidspunktet for indsatsen vurderes ud fra, hvorvidt indsatsen er implementeret i nyere tid (før/efter 1990).

Endelig er det væsentligt at vurdere, hvorvidt indsatsen i et givet studie er møntet en repræsentativ population eller en specifik målgruppe. I en vurdering af modelrelevansen af studiet vil studier, der er populationsrepræsentative, rangere højere end andre på denne subfaktor, da modellen i udgangspunktet baseres på almenområdet. Et eksempel på en målgruppe, der klassificeres som 'høj grad af selektion', er indsatser målrettet udsatte børn og unge.

For at kunne godtgøre et studies modelrelevans er det desuden nødvendig at konkretisere i hvor høj grad indsats, outcomes og klassetrin, der evalueres i studiet, er *sammenlignelige* med modellens. Dette vurderes ved brug af tre subfaktorer:

3.1 Indsatssammenlignelighed

3.2 Outcome-sammenlignelighed

3.2 Tidshorisont

For hvert studie vil der være tale om tre sammenlignelighedsmål for indsatser: 1) samme indsats i model og studie, 2) indsats i studiet vurderes i nogen grad at overlappende med modelindsatsen og 3) andet. På samme vis kan outcome-relevans vurderes.

Med tidshorisont menes her, hvorvidt indsatsen er evalueret på baggrund af outcomes målt op til og med 1 år, efter indsatsen er afsluttet eller på et senere tidspunkt. Dette er en væsentlig faktor, da effekten af en indsats ofte vil ændre sig over tid.

Da modellen har behov for klassetrinspecifikke estimater foretages en registrering af studiets populationsaldersgruppe. Afstanden fra studiets klassetrin til det konkrete moderelevante klassetrin kan herved vurderes.

I tilfælde, hvor et studie indeholder flere relevante indsatser, outcomes og målgrupper, vil disse indgå som separate indgange i datasættet.

Bilag 3 indeholder en oversigt over faktorer, subfaktorer og klassificeringsniveauer.

2.2 Fra klassificering til vægte

Den overfor beskrevne klassificering af studier giver et flerdimensionelt mål for studiets modelrelevans. Dette er imidlertid ikke tilstrækkeligt for at foretage en simpel rangering af studier for dernæst at sammenfatte viden på tværs af studier. Det vil derfor være nødvendigt at kollapse det flerdimensionelle mål i en indikator. Dette gøres via vægte. Denne *relevansindikator* konstrueres konkret som et tal mellem 0 og 1, hvor værdien 1 gives til et studie, der møder alle modelrelevanskriterier til fulde for en given modelindsats og et givet model-outcome.

Hvert studie bliver tilknyttet en relevansindikator for hver kombination af model-outcome, indsats og klassetrin. Relevansindikatoren er større end 0 i de tilfælde, hvor studiet bidrager med relevant viden. Eksempelvis vil et studie, der indeholder estimater på sammenhængen mellem indsatsen tovoksenordningen og outcomes-testresultater i dansk, læsning, indgå med relevansindikator større end 0 i dette tilfælde og med 0 for alle andre outcomes, indsatser. Størrelsen af relevansindikatoren (R) for hvert outcome vil afhænge af studiets klassificering med hensyn til kvalitet, kontekst og sammenlignelighed. For at opbygge en endimensional relevansindikator for hver indsats, outcome og klassetrins kombination bliver det imidlertid også nødvendigt at vægte på tværs af subfaktorer. VIVE foreslår at dette gøres ud fra følgende *to principper*:

8. Subfaktorer vægter ligeligt inden for hver hovedfaktor
9. Hovedfaktorer vægter ligeligt i relevansindikatoren.

Beregningen af relevansindikatoren på baggrund af de to principper er beskrevet nedenfor:

$$R = V_{kvalitet} * V_{kontekst} * V_{sammenlignelighed} * V_{klassetrin}$$

Hvor:

$$V_{kvalitet} = v_{udgivelsesniveau} * v_{empirisk\ metode} * v_{datagrundlag}$$

$$V_{kontekst} = v_{geografisk\ område} * v_{tidspunkt\ for\ indsats} * v_{population\ repræsenterativ}$$

$$V_{sammenlignelig} = v_{indsats} * v_{outcome} * v_{tidshorizont}$$

Relevansindikatoren opbygges af indikatorer, der for hver subfaktor, som fx geografisk område, giver værdien 1 til danske studier og en mindre værdi til

andre studier med mindre modelrelevans. For at gøre dette muligt er det nødvendigt at omsætte hver subfaktorklassificering til numeriske værdier mellem 0 og 1. Følgende algoritme benyttes:

- Et studie, der opfylder kriteriet for subfaktoren til fulde, får vægt 1
- Et studie, der kun opfylder det lavest rangerede kriterium for subfaktoren (minimumskriteriet), får vægt 0,5
- Studier der opfylder kriteriet for subfaktoren på niveau mellem minimumskriterie og fuldt opfyldt, får vægt baseret på en konkav aftagende funktion mellem maksimum (1) og minimum (0,5). Se Bilag 2 for de konkrete niveauer.⁴

Hvert studie tillades kun at indgå i et vægtet estimat for en given kombination af indsats, outcome og klassetrin med en gang. Det er dog ikke et krav, at et studie er beregnet på det eksakte klassetrin, hvor estimatet beregnes. Der trækkes således viden på tværs af klassetrin sådan, at et studie, der beregner en effekt på 5. klasse, også kan indgå i beregningen af det vægtede estimat for eksempelvis 7. klasse. Afstanden mellem det relevante klassetrin og studiets klassetrin indgår i vægtningen. Specifikt indgår klassetrin som en selvstændig subfaktor i vægtningsberegningen med følgende vægt: $\frac{1}{0,25+|x-k|+1}$, hvor x betegner det relevante klassetrin, og k betegner studiets klassetrin. Dette betyder eksempelvis, at et studie på klassetrinnet over eller under det ønskede for beregning, dvs. med en klasseafstand ($|x - k|$) på 1, vil modtage vægten 0,8 (se afsnit 2.5.1 for en diskussion af robustheden). For studier, der beregner estimater for flere klassetrin, imødekommes kravet om, at hvert studie indgår i beregningen af hvert estimat maksimalt en gang ved at vælge estimatet med den korteste klasseafstand.

2.3 Vægtede estimater

På baggrund af hvert studies relevansindikator beregnes for hver kombination af indsats, outcome og klassetrin et vægtet estimat og et evidensniveau.

Det vægtede effektestimater for hver kombination af indsats, outcome og klassetrin beregnes som et vægtet gennemsnit på tværs af estimater fra studier (β), hvor vægtene er relevansindikatoren for hvert studie (R). Det vægtede effektestimater ($\bar{\beta}_x$) for en given kombination af indsats, outcome og klassetrin (x) er givet ved:

⁴ Konkret benyttes en logaritmefunktion. Det vil sige, ved to niveauer mellem maksimum og minimum vil den funktionelle form være: $y = 0,3607\ln(x) + 0,5$ og for et niveau vil den funktionelle form være: $y = 0,4551\ln(x) + 0,5$.

$$\bar{\beta}_x = \frac{\sum_{i=1}^N \beta^i * R_x^i}{\bar{R}_x}$$

2.4 Evidensniveau

Evidensniveauet for hver kombination af indsats, outcome og klassetrin beregnes ved brug af en vidensniveauindikator og et mål for estimaternes samstemmighed. Vidensniveauindikatoren (\bar{R}_x) beregnes som summen over relevansindikatorer. Det vil sige:

$$\bar{R}_x = \sum_{i=1}^N R_x^i,$$

Vidensniveauindikatoren måler derved mængden af relevant viden for det konkrete vægtede estimat.

For at tage hånd om situationer, hvor der findes en række studier med høj relevansindikator, men hvor estimaterne er af modstridende fortegn, medtages afslutningsvis begrebet *samstemmighed*. Samstemmigheden (N_x) beregnes som den største andel af den samlede vidensmængde med samme fortegn og er dermed et mål for spredningen i estimaterne. Det vil sige:

$$N_x = \frac{\max(\bar{R}_x^P, \bar{R}_x^0, \bar{R}_x^N)}{\bar{R}_x}, \text{ hvor } \bar{R}_x^j = \sum_{i=1}^N R^i * I_x^i * I_j^i,$$

$$j \in \{\text{Posivt estimat}, \mathbf{0} \text{ eller insignifikant estimat}, \text{Negativt estimat}\}$$

Det samlede evidensniveau for en given kombination af indsats, outcome og klassetrin beregnes ved følgende algoritme⁵:

Evidensniveaualgoritme:

- Hvis $\bar{R}_x < 0,75$ eller $N_x < 0,6$ sættes evidensniveauet til **Usikkert**
- Hvis $\bar{R}_x \in [0.75; 1.75[$ og $N_x \geq 0,6$ sættes evidensniveauet til **Indikation**

⁵ Løst baseret på algoritmen i jobeffekter.dk. Da antallet af faktorer og subfaktorer samt vægte er forskellige fra algoritmen i jobeffekter.dk, kan kriterierne og niveauerne ikke direkte sammenlignes. Da antallet af faktorer er højere end i jobeffekter.dk, er det naturligt for at opnå et evidensniveau, at det i nogen grad er sammenligneligt, at niveauer for Vidensniveauindikatoren er lavere i modellen end i jobeffekter.dk. Niveauer og kriterier for modellen er valgt på baggrund af sensitivitetanalyser. En yderligere forskel er, at målet for samstemmighed er beregnet forskelligt i de to modeller.

- Hvis $(\bar{R}_x > 1.75$ og $N_x \in [0,6; 0,65[)$ eller $(\bar{R}_x \in [1.75; 2.75[$ og $N_x \geq 0,65)$ sættes evidensniveauet til **Moderat evidens**
- Hvis $\bar{R}_x \geq 2.75$ og $N_x \geq 0,65$ sættes evidensniveauet til **Stærk evidens**.

Vægtningsmodulets forskellige komponenter er sammenfattet i Bilag 3.

2.5 Robusthedsanalyse

Sammenvejning af videnskabelig evidens fra flere forskellige kilder er i sagens natur afhængig af de forskellige valg, der er taget under udarbejdelsen af vægtningsmodulet. I det følgende behandles robustheden af både de vægtede effektestimater, der indgår i modellen, og det resulterende evidensniveau for modellens indsatser. Robusthedsanalysen er foretaget på baggrund af det foreliggende datagrundlag og de implementerede indsatser. Hvis modellem udvides væsentligt med ny data eller med nye indsatser, bør robustheden vurderes på ny.

2.5.1 Robusthed af estimater

I dette afsnit beskrives, hvorledes modelestimaterne påvirkes af ændringer i vægtningsmodulets parametre. Der foretages to analyser⁶:

10. Ændringer af estimater ved ændringer i parameterværdien for klasstrinsafstandsafskrivning
11. Ændringer af estimater ved ændringer i funktional form for relevansafskrivning.

Ændringer af estimater ved ændringer i parameterværdien for klasstrinsafstandsafskrivning

Ved klasstrinsafskrivning forstås, i hvor høj grad estimater fra forskningslitteraturen for et givet klassetrin (x) tillades at påvirke modelestimatet for et andet klassetrin (kl). Denne afskrivning følger følgende formel $\frac{1}{p^{*|x-kl|+1}}$, hvor p i nævneren er en parameterværdi, hvis størrelse angiver graden af afskrivning på tværs af klassetrin. I vægtningsmodulet sættes denne parameter til 0,25. I det følge præsenteres resultater, hvor værdien varieres fra 0 til 1. Når klasstrinsafskrivningen antager værdien 0, vil effektestimater have samme vægt på

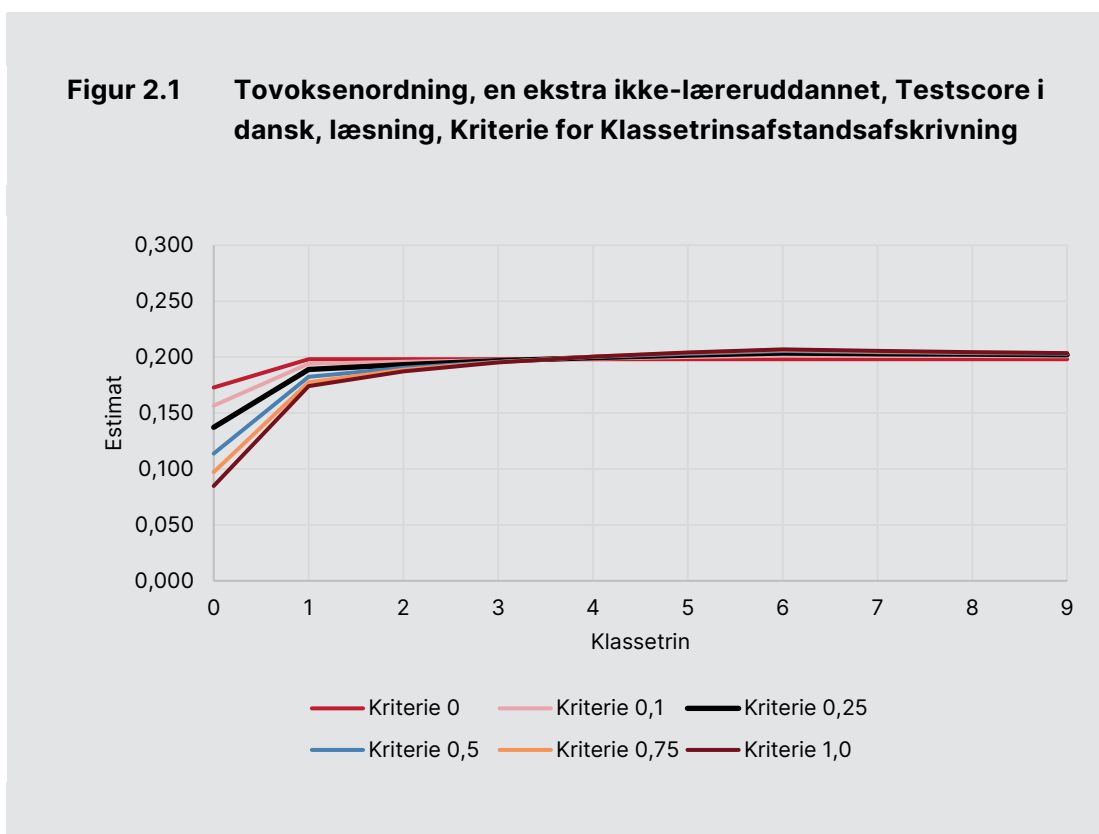
⁶ Analyserne er foretaget for indsatserne Tovoksenordning og Klassekvotient og med udgangspunkt i den indsamlede evidens. Robustheden kan ændres ved implementeringen af nye indsatser og ved tilføjelse af ny evidens.

tværs af klassetrin uanset klassetrinnet for indsatsen i litteraturen. Når værdien er 1, vil effektstørrelser målt på andre klassetrin end det modelrelevante straffes relativt hårdere i forhold til effektestimater fra samme klassetrin, desto større afvigelsen er mellem indsatsklassetrinnet og modellens klassetrin. Figur 2.1 og Figur 2.2 viser variationen i estimater ved at varierer klassetrinsafskrivningen for hhv.:

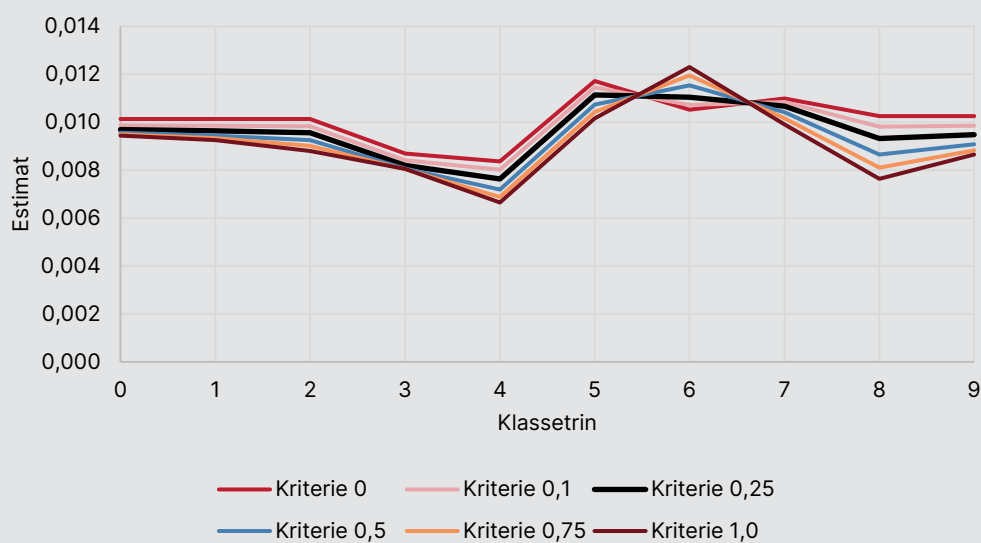
12. Tovoksenordning, en ekstra ikke-læreruddannet, testscore i dansk, læsning
13. Klassekvotient, testscore i matematik.

De to cases er valgt, da 1) repræsenterer et tilfælde med få studier, og 2) repræsenterer et tilfælde med mange studier og på mange klassetrin, men skiftende fortegn på estimater.

Samlet viser Figur 2.1 og Figur 2.2, at modelestimaterne kun i mindre grad er sensitive ved ændringer i denne parameter, når resten af modellen holdes konstant.



Figur 2.2 Klassekvotient, Testscore i Matematik, Kriterie for Klassetrinsafstandsafskrivning



Ændringer af estimer ved ændringer i funktionel form for relevansafskrivning

Relevansafskrivning betegner grad af vægt, der lægges på studier, der ikke fuldt ud opfylder klassifikationskriterierne. Der foretages fem robusthedsanalyser

1. Lineær: Der antages en lineær relevansafskrivning.⁷ I analyserne fastholdes, at studier, der har et minimum af relevans, tager vægten 0,5, og studier, der opfylder klassifikationskriteriet til fulde, tager værdien 1.
2. EXP: Der antages en eksponentiel relevansafskrivning.⁸ I analyserne fastholdes, at studier, der har et minimum af relevans, tager vægten 0,5, og studier, der opfylder klassifikationskriteriet til fulde, tager værdien 1.
3. 0,75-vægte: Studier, der opfylder klassifikationskriteriet til fulde, tager værdien 1. Andre studier tager værdien 0,75.

⁷ Det vil sige, ved to niveauer mellem maksimum og minimum vil den funktionelle form være: $y = 0,1667x + 0,3333$ og for et niveau vil den funktionelle form være: $y = 0,25x + 0,25$.

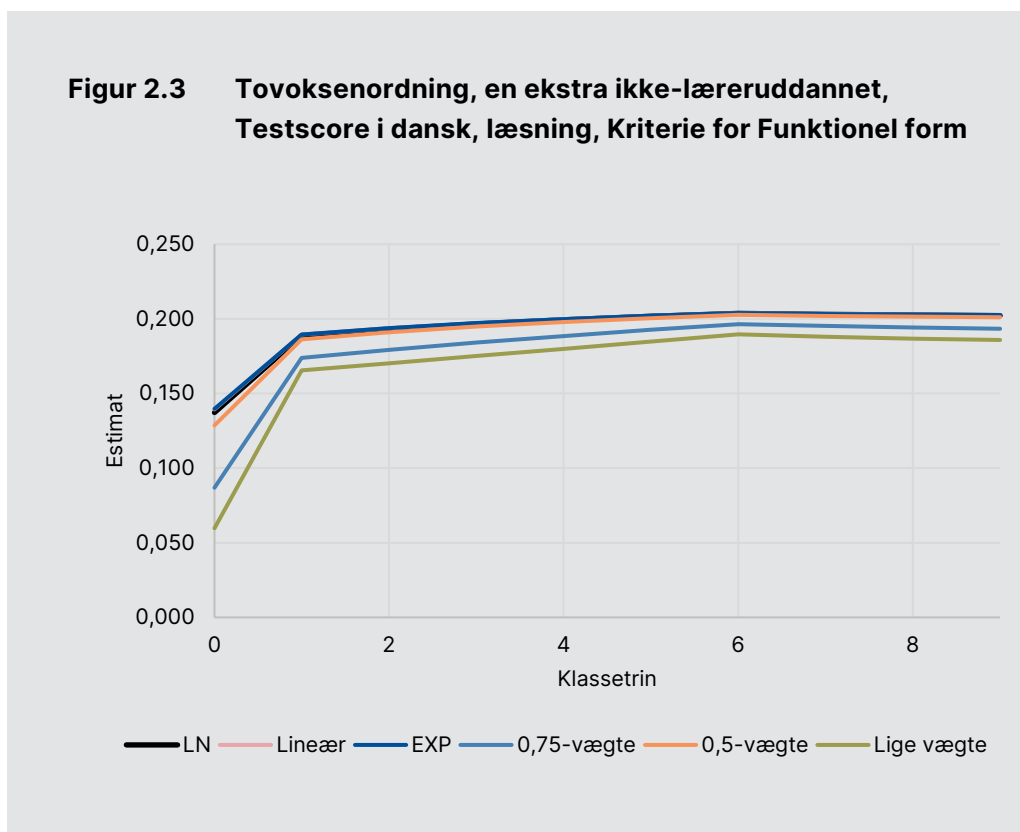
⁸ Det vil sige, ved to niveauer mellem maksimum og minimum vil den funktionelle form være: $y = 0,3969e^{0,231x}$ og for et niveau vil den funktionelle form være: $y = 0,3536e^{0,3466x}$.

4. 0,50-vægte: Studier, der opfylder klassifikationskriteriet til fulde, tager værdien 1. Andre studier tager værdien 0,50.
5. Lige vægte: Alle studier tager vægten 1.

Disse sammenholdes med modelspecifikationen:

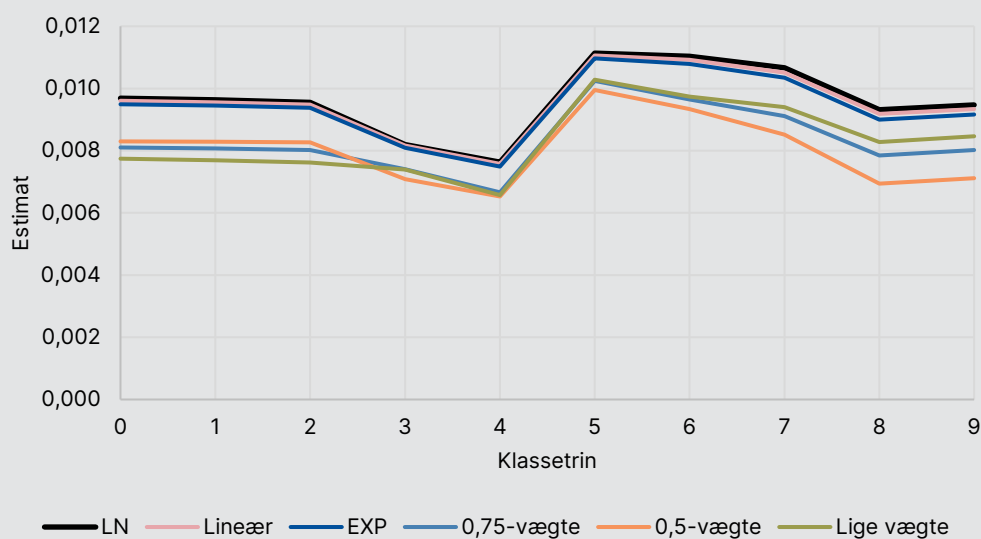
- *LN*: Der antages en logaritmisk relevansafskrivning.⁹ I analyserne fastholdes, at studier, der har et minimum af relevans, tager vægten 0,5, og studier, der opfylder klassifikationskriteriet til fulde, tager værdien 1.

Figur 2.3 og Figur 2.4 viser resultater af disse analyser for de to cases: Tovoksenordning, en ekstra ikke-læreruddannet, testscore i dansk, læsning og Klassekvotient, testscore i matematik. Resultaterne viser, at modelestimerne kun i mindre grad er sensitive til ændringer i relevansafskrivningen.



⁹ Det vil sige, ved to niveauer mellem maksimum og minimum vil den funktionelle form være: $y = 0.3607\ln(x) + 0.5$ og for et niveau vil den funktionelle form være: $y = 0.4551\ln(x) + 0.5$.

Figur 2.4 Klassekvotient, Testscore i Matematik, Kriterie for Funktional form



2.5.2 Robusthed af evidensniveauer

I dette afsnit beskrives, hvorledes evidensniveauet påvirkes af ændringer i vægtningsmodulets parametre. Evidensniveauet fastlægges på baggrund af algoritmen beskrevet i afsnit 2.4. De to væsentligste modelementer, der har betydning for algoritmen, er vidensmængden og samstemmigheden af estimerne. For at undersøge robustheden af vægtningsmodulets evidensniveau foretages to analyser:

6. Robusthed ved ændring i krav til vidensmængden
7. Robusthed ved ændring i krav til samstemmigheden.

Robusthed ved ændring i krav til vidensmængden

Robustheden ved ændring i krav til vidensmængden vurderes ved at beregne modelversioner med varierende krav til vidensmængden. Dette gøres konkret ved at variere cut-offs (vm) for vidensindikatoren i evidensniveaualgoritmen. Evidensniveaualgoritmen er nærmere beskrevet i afsnit 2.4, men gengivet nedenfor.

- Hvis $\bar{R}_x < vm$ eller $N_x < 0,6$ sættes evidensniveauet til **Usikkert**
- Hvis $\bar{R}_x \in [vm; 1 + vm[$ og $N_x \geq 0,6$ sættes evidensniveauet til **Indikation**

- Hvis $(\bar{R}_x > 1 + vm$ og $N_x \in [0,6; 0,65[)$ eller $(\bar{R}_x \in [1 + vm; 2 + vm[$ og $N_x \geq 0,65)$ sættes evidensniveauet til **Moderat evidens**
- Hvis $\bar{R}_x \geq 2 + vm$ og $N_x \geq 0,65$ sættes evidensniveauet til **Stærk evidens**.

Tabel 2.2 viser resultater for evidensniveau for indsatserne Tovoksenordning, en ekstra ikke-læreruddannet og Klassekvotient med link outcomes hhv. test-score i dansk, læsning og testscore i matematik. Resultaterne præsenteres for hvert klassetrin og for variationer af vidensindikator cut-offs.

Resultaterne viser, at et øget krav til evidensniveau (dvs. højere cut-offs) medfører lavere evidensniveauer overordnet set. Dette gør sig i høj grad gældende for indsats-outcome-kombinationen Tovoksenordning, en ekstra ikke-læreruddannet, testscore i dansk, læsning, hvor der er en lille mængde af viden tilgængelig, og i mindre grad gældende for indsats-outcome-kombinationen Klassekvotient, testscore i matematik, hvor der er mere viden.

Tabel 2.2 Robusthed ved ændring i krav til vidensmængden

Vidensindikator cut-offs					
Link outcome	Klassetrin	vm = 0,50	vm = 0,75	vm = 1,00	vm = 1,25
Tovoksenordning, en ekstra ikke-læreruddannet					
Testscore i dansk, læsning	0	Indikation	Usikker	Usikker	Usikker
Testscore i dansk, læsning	1	Indikation	Usikker	Usikker	Usikker
Testscore i dansk, læsning	2	Indikation	Usikker	Usikker	Usikker
Testscore i dansk, læsning	3	Indikation	Usikker	Usikker	Usikker
Testscore i dansk, læsning	4	Indikation	Usikker	Usikker	Usikker
Testscore i dansk, læsning	5	Indikation	Indikation	Usikker	Usikker
Testscore i dansk, læsning	6	Indikation	Indikation	Usikker	Usikker
Testscore i dansk, læsning	7	Indikation	Indikation	Usikker	Usikker
Testscore i dansk, læsning	8	Indikation	Usikker	Usikker	Usikker
Testscore i dansk, læsning	9	Indikation	Usikker	Usikker	Usikker
Klassekvotient					

Vidensindikator cut-offs					
Testscore i matematik	0	Moderat evidens	Moderat evidens	Indikation	Indikation
Testscore i matematik	1	Moderat evidens	Moderat evidens	Moderat evidens	Indikation
Testscore i matematik	2	Moderat evidens	Moderat evidens	Moderat evidens	Moderat evidens
Testscore i matematik	3	Usikker	Usikker	Usikker	Usikker
Testscore i matematik	4	Usikker	Usikker	Usikker	Usikker
Testscore i matematik	5	Stærk evidens	Stærk evidens	Stærk evidens	Stærk evidens
Testscore i matematik	6	Stærk evidens	Stærk evidens	Stærk evidens	Stærk evidens
Testscore i matematik	7	Moderat evidens	Moderat evidens	Moderat evidens	Moderat evidens
Testscore i matematik	8	Usikker	Usikker	Usikker	Usikker
Testscore i matematik	9	Usikker	Usikker	Usikker	Usikker

Robusthed ved ændring i krav til samstemmigheden

Robustheden ved ændring i krav til samstemmigheden vurderes ved at beregne modelversioner med varierende krav til samstemmigheden af forskningslitteraturen. Dette gøres konkret ved at variere cut-off (nv) for samstemmigheden i evidensniveaualgoritmen. Samstemmigheden (N_x) beregnes som den største andel af den samlede vidensmængde med samme fortegn (jf. afsnit 2.4). Evidensniveaualgoritmen er gengivet nedenfor.

- Hvis $\bar{R}_x < 0,75$ eller $N_x < nv$ sættes evidensniveauet til **Usikkert**
- Hvis $\bar{R}_x \in [0,75; 1,75[$ og $N_x \geq nv$ sættes evidensniveauet til **Indikation**
- Hvis ($\bar{R}_x > 1,75$ og $N_x \in [nv; nv + 0,05[$) eller ($\bar{R}_x \in [1,75; 2,75[$ og $N_x \geq nv + 0,05$) sættes evidensniveauet til **Moderat evidens**
- Hvis $\bar{R}_x \geq 2,75$ og $N_x \geq nv + 0,05$ sættes evidensniveauet til **Stærk evidens**.

Tabel 2.3 viser resultater for evidensniveau for indsatserne Tovoksenordning, en ekstra ikke-læreruddannet og Klassekvotient med link outcomes hhv. testscore i dansk, læsning og testscore i matematik. Resultaterne præsenteres for hvert klassetrin og for variationer af cut-off for samstemmighed.

Resultaterne viser, at et øget krav til evidensniveau (dvs. højere cut-off) medfører lavere evidensniveauer for indsats-outcome-kombinationen Klassekvotient, testscore i matematik, hvor der er væsentlig spredning i samstemmigheden af viden. Ændring i kravet til samstemmighed af viden har derimod mindre

betydning for evidensniveauet for indsats-outcome-kombinationen Tovoksenordning, en ekstra ikke-læreruddannet, testscore i dansk, læsning, hvor den lille mængde viden udviser en højere grad af samstemmighed.

Tabel 2.3 Robusthed ved ændring i krav til samstemmigheden

Nettovidensmængde cut-offs					
Link outcome	Klasse-trin	nv = 0,5	nv = 0,6	nv = 0,7	nv = 0,8
Tovoksenordning, en ekstra ikke-læreruddannet					
Testscore i dansk, læsning	0	Usikker	Usikker	Usikker	Usikker
Testscore i dansk, læsning	1	Usikker	Usikker	Usikker	Usikker
Testscore i dansk, læsning	2	Usikker	Usikker	Usikker	Usikker
Testscore i dansk, læsning	3	Usikker	Usikker	Usikker	Usikker
Testscore i dansk, læsning	4	Usikker	Usikker	Usikker	Usikker
Testscore i dansk, læsning	5	Indikation	Indikation	Indikation	Indikation
Testscore i dansk, læsning	6	Indikation	Indikation	Indikation	Indikation
Testscore i dansk, læsning	7	Indikation	Indikation	Indikation	Indikation
Testscore i dansk, læsning	8	Usikker	Usikker	Usikker	Usikker
Testscore i dansk, læsning	9	Usikker	Usikker	Usikker	Usikker
Klassekvotient					
Testscore i matematik	0	Moderat evidens	Moderat evidens	Usikker	Usikker
Testscore i matematik	1	Moderat evidens	Moderat evidens	Usikker	Usikker
Testscore i matematik	2	Moderat evidens	Moderat evidens	Usikker	Usikker
Testscore i matematik	3	Moderat evidens	Usikker	Usikker	Usikker
Testscore i matematik	4	Stærk evidens	Usikker	Usikker	Usikker
Testscore i matematik	5	Stærk evidens	Stærk evidens	Moderat evidens	Usikker
Testscore i matematik	6	Stærk evidens	Stærk evidens	Moderat evidens	Usikker
Testscore i matematik	7	Stærk evidens	Moderat evidens	Usikker	Usikker
Testscore i matematik	8	Moderat evidens	Usikker	Usikker	Usikker
Testscore i matematik	9	Moderat evidens	Usikker	Usikker	Usikker

3 Multiple typer af evidens og mekanisme til prioritering

Beregningsmodulet til REFUD udvikles til at kunne håndtere multiple link outcomes og direkte evidens inden for samme indsats.¹⁰ Dette medfører et behov for, at beregningsmodulet skal være i stand til at prioritere mellem multiple typer af evidens. I det følgende beskrives, hvorledes multiple typer af evidens og en mekanisme til prioritering er implementeret i beregningsmodulet.

3.1 Direkte evidens

Direkte evidens betegner det tilfælde, hvor der i forskningslitteraturen findes et eller flere estimater på en effekt af en indsats på outcomes, som anvendes i REFUD som konsekvenser. I det tilfælde er det ikke nødvendigt at bero sig på link outcomes. Modellen udvides til at kunne håndtere sådanne tilfælde. Det er afgørende, at direkte evidens kun tilføjes modellen, hvis det vurderes at have en kvalitet, så den empiriske sammenhæng kan tilskrives en årsagssammenhæng, da denne prototypemodel altid prioriterer direkte evidens over indirekte evidens. I praksis håndteres direkte evidens i beregningsmodulet ved at overskrive linkdata med ettaller og anvende særskilt direkte beta-estimat. Konsekvenser med direkte evidens specificeres i filen `indsatser.json`.

Muligheden for at regne på direkte evidens skaber et behov for, at modellen kan håndtere persistens af disse effekter. Det vil sige, om effekten er aftagende, konstant eller stigende over tid (Lesner et al., 2022). Årsagen til dette er, at modellen skal kunne beregne konsekvenser med en fireårig tidshorisont, men at effektestimater for hvert af de fire år efter indsatsstart ikke forventes tilgængelig i forskningslitteraturen. Der vil derfor være behov for, at modellen tager eksplicit højde for persistens i estimaterne over tid. Det vil sige, i hvor høj grad det forventes, at effekten af indsatsen fastholdes over tid for hver konsekvens. Dette gøres ved at konstruere en særskilt persistensmatrice som inputdata i modellen. Denne beregnes ved hjælp af en statistisk regressionsmodel, hvor der samtidig tages højde for en række baggrundsfaktorer om barnet og barnets initiale forbrug af flere af ydelserne.

¹⁰ Med multiple link outcomes forstås det tilfælde, hvor samme indsats har signifikant effekt på flere forskellige link outcomes. Det kan eksempelvis være en skoleindsats, der både påvirker elevernes læring og trivsel, og hvor modellen således tilbyder effektberegninger på de endelige konsekvenser baseret på effektstørrelserne fra litteraturen på begge. Med direkte evidens forstås det tilfælde, hvor samme indsats har signifikant effekt på et eller flere outcomes, hvor der umiddelbart kan knyttes økonomi, og hvorfor de derfor betragtes som konsekvenser i modellen.

For hver konsekvens ($Y_k, k = 1, \dots, K$) estimeres således følgende regressionsmodel for hvert af indsatsens relevante klassetrin (i det følgende undertrykkes subskription for konsekvensnummer og klassetrin) separat for årene $t = 2, \dots, 4$:

$$Y_{it} = \beta_t^0 + \beta_t^{pers} Y_{i1} + \beta_t^1 X_i + \gamma_t^m + \epsilon_{it},$$

Hvor Y_{i1} er individets (i) konsekvensforbrug i år 1, X_i er en vektor af elevspecifikke baggrundsvARIABLE, herunder initialt forbrug af udvalgte offentlige ydelser til tidspunkt 0 (se også Lesner et al., 2022, Tabel 5.1 for en nærmere beskrivelse af de inkluderede regressorer), og γ^m er kommune fixed-effects. I overensstemmelse med de øvrige estimater, der benyttes i modellen klyngekorrigeres standardfejlene på kommunalt niveau. Persistensmatricen indeholder således i udgangspunktet $\hat{\beta}_2^{pers}, \dots, \hat{\beta}_4^{pers}$ fra ovenstående regressioner, dog med den undtagelse, at estimater, der ikke er signifikante på et 1 %-niveau, tvinges til 0 i praksis. For implementeringen af matricen i beregningsmodulet, sættes persistensen af effekten på konsekvenserne i år 1 ($\hat{\beta}_1^{pers}$) til 1 for alle konsekvenser.

3.2 Multipel evidens og prioritering

Hvis man ønsker at tage højde for flere kanaler, ad hvilke en indsats kan påvirke konsekvenserne, risikerer man dobbelttælling. Hvis et studie eksempelvis peger på en effekt af reduceret klassekvotient på elevernes læringsresultater, og et andet studie eksempelvis peger på en effekt af reduceret klassekvotient på elevernes trivsel, vil det overestimere effekten på konsekvenserne i REFUD, hvis man blot lægger effekterne sammen (såfremt fortegnet er det samme). Årsagen hertil er, at det er sandsynligt, at begge link outcomes (læsning og trivsel) fanger et overlap af effekterne ved indsatsen.

For at undgå dobbelttællingsproblematikken skal der i beregningsmodulet implementeres en mekanisme, som tager hånd om dette. Denne prioriteringsmekanisme implementeres i tråd med beskrivelsen i Jacobsen et al. (2022). Konkret fungerer prioriteringsmekanismen på følgende måde:

Ved hvert link outcome (hvortil der er signifikante betaer) udregnes de samlede økonomiske konsekvenser for alle indsatsrelevante outcomes, og der tilbagediskonteres over de fire år. Der anvendes kun signifikante estimater på betaer og på de økonomiske konsekvenser.

VIVE foreslår, at det endelige resultat beregnes med afsæt i et valg mellem tre forskellige 'beregningsskemaer':¹¹

8. Den mest positive effekt. For hver konsekvens tages den største tilbagediskonterede værdi.
9. Den mest konservative effekt. For hver konsekvens tages den mindste tilbagediskonterede værdi.
10. Gennemsnit af effekter. For hver konsekvens udregnes det simple gennemsnit af alle værdier.

Hvis der er både positive og negative beta-estimer på et område, hvor vi ville forvente positive estimer, anser vi det for at være det mest konservative at vælge den mindste, altså den negative og mest ufavorable værdi, da dette giver den mindste samlede beregnede effekt ved implementering af indsatsen. Beregningsskemaerne udregnes baseret på brugerens input (valg af klassetrin og tidshorisont) og er således beregnings-specifik.

¹¹ Modellen kan udvides med yderligere paradigmer, hvis dette skønnes relevant for brugeren.

4 Ny indsats: Klassekvotient

I det følgende beskrives hvorledes indsatsen Klassekvotient afgrænses og implementeres som indsats i REFUD.

4.1 Helårseffekter

I REFUD betegnes klassekvotient-indsatsen som effekten af at reducere en klassestørrelse med én elev i alle årets undervisningstimer (helårseffekt). Ud fra en linearitetsantagelse ganges denne helårseffekt herefter med brugerens valgte klassestørrelsesreduktion¹².

På baggrund af klassetrinspecifikke effektstørrelser (se vægtningsmodulet i Afsnit 2) benytter modellen estimer omregnet til helårseffekter til beregning af forventede nettogevinster ved indsatsen. De inkluderede effektstørrelser fra forskningslitteraturen spænder overordnet set fra insignifikante til moderate positive effekter af en klassestørrelsesreduktion. Jævnfør Jacobsen et al. (2022) skelnes i modellen ikke mellem empiri baseret på klassestørrelse og klassekvotient i de i vægtningsmodulet inkluderede effektestimater.

4.2 Målgruppe og depreciering

Indsatsen implementeres umiddelbart i den eksisterende modelramme for REFUD, der indeholder folkeskolens normalklasseelever på 0.-9. klassetrin (Lesner et al., 2022)).

Det er i modellen muligt at specificere indsatsdosis ved at angive antallet af elever, som klassestørrelsen reduceres med. Resultatet af beregningerne i modellen baserer sig udelukkende på de i modellen indtastede 'indsatsbørn', dvs. de børn, der modtager den reducerede klassestørrelse (se også Lesner et al., 2022). Modellen tager ikke højde for generelle ligevægtseffekter såsom effekten på børn, der ville have gået i indsatsklasse, hvis der ikke havde været en ændring i klassestørrelsen.

Modellen indeholder i udgangspunktet ingen begrænsninger i forhold til størrelsen på klassestørrelsesreduktionen, der benyttes i beregningerne. Brugeren bør derfor være opmærksom på, at de beregnede indsats effekter kun i mindre

¹² Ved beregning på effekterne af en klassestørrelsesforøgelse indgår helårseffekten blot med modsat fortegn.

grad er gældende for klassestørrelser, der afviger fra klassestørrelserne i de inkluderede studier, hvor den gennemsnitlige størrelse generelt varierer fra 20-25 elever (undtaget herfra er Angrist & Lavy (1999), der er baseret på israelske folkeskoler med op til 40 elever i en klasse).

Brugeren kan ligeledes foretage beregninger på en klassestørrelsesreduktion over flere år. Det vil sige de situationer, hvor de samme elever forventes at opleve den samme reducerede klassestørrelse over flere klassetrin. I det tilfælde er det nødvendigt at forholde sig til, hvilken effekt indsatsen vil have, når den modtages gentagne gange.

Ud fra et princip om konsistens i behandlingen af effektestimater i modellen regner modellen i udgangspunktet med ens effektstørrelser i første og eventuelle efterfølgende indsatsår. Antagelse om lineære ekstrapolation af effektstørrelser over år er ikke empirisk funderet. Idet der er en relativt lille variation i elevers klassestørrelser over klassetrin, er det i den kvasiexperimentielle del af forskningslitteratur svært at udskille effekten af klassestørrelser på tidligere klassetrin fra nuværende klassestørrelser, mens paneldatamodellerne netop benytter sig af denne variation til at estimere klassestørrelseseffekten. Dermed belyser de inkluderede studier ikke direkte effekten af gentagne (fortsatte) reduktioner. Effekten af klassestørrelsesreduktioner er dog relativt vel-dokumenteret for flere niveauer af klasse (fx (Nandrup, 2016), og her fremgår det, at effektstørrelsen ikke nødvendigvis er konstant. Denne både stiger og falder med klassetrinnet, hvorfor det ikke er ligetil at udlede en generel deprecieringsfaktor. I mangel af konkret viden vurderes det derfor mest hensigtsmæssigt fortsat at benytte en lineær ekstrapolation af effektestimaterne.

Hvis brugeren ud fra konkret viden mener, at antagelsen om lineære ekstrapolation af gentagne indsatser ikke er retvisende, er det som ved alle andre parametre i modellen muligt at justere deprecieringen. Det er desuden muligt at justere i begge retninger, således at en indsat kan have både stigende og aftagende margineffekt over indsatsår.

5 Nyt link outcome: Elevtrivsel

I udvidelsen af REFUD-prototypen inkluderes elevtrivsel i forskellige dimensioner som link outcomes. Dette gøres efter ønske fra BUVM i en forståelse af, at ændring i trivsel i skolen og skolelivet generelt kan være et væsentligt udfald af indsatser på børne- og undervisningsområdet, og skal ses som en tilføjelse til de eksisterende link outcomes i REFUD-prototypen, jf. Lesner et al. (2022). Specifikt tilføjes fem særskilte link outcomes baseret på Den Nationale Trivselsmåling til beregning af indsatseffekter med udgangspunkt i elevers trivsel i folkeskolen på modellens konsekvenser: faglig trivsel, social trivsel, støtte og inspiration, ro og orden, og slutteligt et mål for generel trivsel (se også Styrelsen for It og Læring, 2017). Disse trivselsmål baserer sig på spørgerammen til 4. til 9. klasse og er derfor kun tilgængelige for disse klassetrin. De indgår på samme vis som tidligere link outcomes i estimering af lokale standardafvigelses på link outcomes og danner grundlaget for den betingede sammenhæng mellem link outcomes og konsekvensforbrug¹³. De nye link outcomes er beskrevet i Tabel 5.1 og 5.2.

Tabel 5.1 Tilgængelige link outcomes for elevtrivsel i REFUD

Link outcome	Kilde	Tilgængelige klassetrin	Registerbehandling	Link-nummer
Generel trivsel	Den Nationale Trivselsmåling	4.-9. klasse	Indikatoren for generel trivsel er et samlet mål for trivsel konstrueret ved et simpelt gennemsnit af de 29 spørgsmål, der bruges til dannelse af de 4 primære indikatorer. For elever, der har svaret på færre end 15 spørgsmål, er der ikke dannet en værdi til indikatoren.	3
Faglig trivsel	Den Nationale Trivselsmåling	4.-9. klasse	Indikatoren for faglig trivsel er konstrueret ved et simpelt gennemsnit af 8 spørgsmål. For elever, der har svaret på færre end 4 spørgsmål, er der ikke dannet en værdi til indikatoren.	4
Social trivsel	Den Nationale Trivselsmåling	4.-9. klasse	Indikatoren for social trivsel er konstrueret ved et simpelt gennemsnit af 10 spørgsmål. For elever, der har svaret på færre end 5 spørgsmål, er der ikke dannet en værdi til indikatoren.	5
Støtte og inspiration	Den Nationale Trivselsmåling	4.-9. klasse	Indikatoren for støtte og inspiration er konstrueret ved et simpelt gennemsnit af 7 spørgsmål. For elever, der har svaret på færre end 4 spørgsmål, er	6

¹³ Dog med den undtagelse, at trivselsmålene ikke ekstrapoleres til, hvor de er utilgængelige (0.-3. klassetrin) på samme vis som link outcomes baseret på de nationale test.

Link outcome	Kilde	Tilgængelige klassetrin	Registerbehandling	Link-nummer
			der ikke dannet en værdi til indikatoren.	
Ro og orden	Den Nationale Trivselsmåling	4.-9. klasse	Indikatoren for ro og orden er konstrueret ved et simpelt gennemsnit af 4 spørgsmål. For elever, der har svaret på færre end 2 spørgsmål, er der ikke dannet en værdi til indikatoren.	7

Anm.: Tabellen angiver REFUDs tilgængelige link outcomes for trivsel, hvordan de er konstrueret, samt for hvilke klassetrin disse er relevante. Dannelse af trivselsindikatorerne er baseret på data fra Den Nationale Trivselsmåling fra STIL. Se tabel 5.2 for inkluderede spørgsmål.

Note: Alle link outcomes er standardiseret med middelværdi 0 og standardafvigelse 1 inden for test-år og klassetrin.

5.1 Konstruktion af link outcomes

Efter aftale med Børne- og Undervisningsministeriet er konstruktionen af de nye link outcomes baseret på trivselsindikatorer i metodenotat af 26.10.2020 af BUVM STIL (Styrelsen for It og Læring, 2017). Her anvendes en faktoranalyse på de 40 spørgsmål i spørgerammen til Den Nationale Trivselsmåling for 4.-9. klasse til at undersøge de underliggende faktorer, som spørgsmålene dækker over. Der er udelukkende udarbejdet indikatorer i relation til spørgerammen, der er anvendt af eleverne i 4.-9. klasse, og der dannes derfor ikke trivselsindikatorer for 0.-3. klasse. Baseret på denne fremgangsmåde identificeres 4 relevante faktorer ud fra 29 spørgsmål, som danner grundlag for konstruktionen af de 4 trivselsindikatorer for 4.-9. klasse. Derudover dannes et samlet trivselsmål for den generelle elevtrivsel baseret på de 4 trivselsfaktorer. For yderligere detaljer om faktoranalysen bag trivselsindikatorerne, se Styrelsen for It og Læring (2017).

I konstruktionen af link outcomes er først de 4 trivselsmål dannet ud fra et simpelt gennemsnit baseret på de i faktoranalysen udvalgte spørgsmål, se Tabel 5.2 (Styrelsen for It og Læring, 2017). Den samlede trivselsindikator for generel trivsel er dannet ud fra et simpelt gennemsnit af alle 29 spørgsmål brugt til dannelse af de 4 trivselsfaktorer. Eleven indgår kun med en værdi for trivselsmålet, hvis denne har besvaret mindst halvdelen af spørgsmålene i den relevante faktor. Efterfølgende er hvert af de fem trivselsmål standardiseret til middelværdi 0 og standardafvigelse 1 inden for klassetrin og testår. Denne standardisering tillader, at trivselsindikatorerne på tværs af klassetrin og år er umiddelbart sammenlignelige. De standardiserede trivselsindikatorer tilføjes ligesom de eksisterende link outcomes baseret på de nationale test til elevrammen ved hjælp af elev-id og skoleår. Herefter beregnes standardafvigelser

for hvert klassetrin og kommune samt betingede sammenhænge med de endelige konsekvenser for hvert klassetrin.¹⁴

Tabel 5.2 Spørgsmål til konstruktion af trivselsindikatorer

Trivselsindikator	Spørgsmål
Faglig trivsel	Hvad synes dine lærere om dine fremskridt i skolen? Lykkes det for dig at lære dét, du gerne vil i skolen? Hvor tit kan du finde en løsning på problemer, bare du prøver hårdt nok? Hvor tit kan du klare det, du sætter dig for? Kan du koncentrere dig i timerne? Jeg klarer mig godt fagligt i skolen. Jeg gør gode faglige fremskridt i skolen. Hvis jeg bliver forstyrret i undervisningen, kan jeg hurtigt koncentrere mig igen.
Social trivsel	Er du glad for din skole? Er du glad for din klasse? Føler du dig ensom? Jeg føler, at jeg hører til på min skole. Jeg kan godt lide pauserne i skolen. De fleste af eleverne i min klasse er venlige og hjælpsomme. Andre elever accepterer mig, som jeg er. Er du blevet mobbet i dette skoleår? Er du bange for at blive til grin? Hvor ofte føler du dig tryk i skolen?
Støtte og inspiration	Hjælper dine lærere dig med at lære på måder, som virker godt? Undervisningen giver mig lyst til at lære mere. Lærerne er gode til at støtte mig og hjælpe mig i skolen. Er du og dine klassekammerater med til at bestemme, hvad I skal arbejde med i klassen? Lærerne sørger for, at elevernes ideer bliver brugt i undervisningen. Er undervisningen kedelig? Er undervisningen spændende?
Ro og orden	Hvis der er larm i klassen, kan lærerne hurtigt få skabt ro. Møder dine lærere præcist til undervisningen? Er det let at høre, hvad læreren siger i timerne? Er det let at høre, hvad de andre elever siger i timerne?

Anm.: Tabellen angiver formuleringen af de spørgsmål, der udgør de fire trivselsmål for 4.-9. klasse. Det femte og overordnede trivselsmål Generel trivsel indeholder spørgsmålene fra hvert af de fire trivselsdelmål.

Kilde: (Styrelsen for It og Læring, 2017)

5.2 Datatilgængelighed

Trivselsindikatorerne er dannet på baggrund af data fra Den Nationale Trivselsmåling, som er tilgængeligt fra skoleåret 2014/2015 til 2019/2020. For at

¹⁴ For 0.-3. klasse er disse estimater sat til 0, da link outcomes ikke er dannet for disse klassetrin.

undgå eventuelle påvirkninger af estimer forårsaget af skolenedlukningen i forbindelse med covid-19-pandemien i foråret 2020, ekskluderes trivselsmålingerne for 2019/2020 i modellen. Dermed benyttes trivselsindikatorerne som link outcomes i modelberegninger i perioden 2014/2015-2018/2019. Dette er konsistent med datagrundlaget for de tidligere link outcomes, jf. Lesner et al. (2022).

Litteratur

Angrist, J. D., & Lavy, V. (1999). Using maimonides' rule to estimate the effect of class size on scholastic achievement. *The Quarterly Journal of Economics*, 114(2), 533-575.

Jacobsen, R. H., Andreasen, A. G., Keilow, M., Krassel, K. F., Lesner, R. V., Mortensen, N. P. Verner, M. (2022). *Afdækning af muligheder for udbredelse af SØM til indsatser på børne- og undervisningsområdet*. København: VIVE – Det Nationale Forsknings- og Analysecenter for Velfærd.

Lesner, R. V., Krassel, K. F., & Nandrup, A. B. (2022). *Udvikling af regnemodel for uddannelsesinvesteringer (REFUD) - Teknisk dokumentationsrapport*. København: VIVE – Det Nationale Forsknings- og Analysecenter for Velfærd.

Lesner, R. V., & Nandrup, A. B. (2022). *Protokol for indskrivning i REFUD vægtningsdatabase*. København: VIVE – Det Nationale Forsknings- og Analysecenter for Velfærd.

Nandrup, A. B. (2016). Do class size effects differ across grades? *Education Economics*, 24(1), 83-95.

Styrelsen for It og Læring. (2017). *Metodenotat: Beregning af indikatorer i den nationale trivselsmåling i folkeskolen trivselsmåling*. København: Børne- og Undervisningsministeriet, Styrelsen for It og Læring.

Bilag 1 Ordforklaringer

β^{ab}	Se indirekte evidens.
β^{ac}	Se direkte evidens.
β^{bc}	Se indirekte evidens.
Baggrundvariabel, baggrundskaraktetika	Oplysninger, der kontrolleres for ved estimation af betingede korrelationer mellem link outcomes og konsekvenser.
Baseline	Tidspunkt for måling af variable. Enten $t = 0$ eller $t = 1$. Se herunder for hvert klassetrin og hver kommune.
Baseline-konsekvensforbrug (individet)	Individets baseline-konsekvensforbrug benyttes som regressor i estimationen af den betingede korrelation mellem link outcome og konsekvens og måles til $t = 0$ (før første indsatsår).
Baseline-standardafvigelse	Baseline-standardafvigelser angiver (lokale) kommunale standardafvigelser på henholdsvis link outcomes og konsekvenser på hvert af analysepopulationens klassetrin og dermed til tid $t = 1$.
Betinget korrelation	Lineær statistisk sammenhæng mellem to variable, hvor der kontrolleres for en eller flere baggrundsvareable. Måles for hvert klassetrin.
Default konsekvenser	Konsekvenser, som med rimelighed kan forventes at blive påvirket af en generel skolebaseret indsats og derfor er valgt på forhånd for brugeren i beregningsmodulet.
Depreciering	Et mål, der beskriver, hvor meget en indsats' effektivitet aftager ved gentagelse. Antages i udgangspunktet at være 0 i mangel på empiri.
Direkte evidens	Det tilfælde, hvor virkningen af en indsats er målt direkte på en eller flere konsekvenser.
Indirekte evidens	Det tilfælde, hvor virkningen af en indsats er målt via et link outcome. Sammenhængen mellem link outcome og konsekvens kan hentes i litteraturen eller estimeres baseret på de administrative registre hos Danmarks Statistik.
Indsats	En handling, der ændrer status quo og giver anledning til at forvente økonomiske konsekvenser over en 4-årig periode.
Indsatsvariant	Forskellige, men dog sammenlignelige udformninger af en given indsats.
Ekstensiv margin	Indsatsens omfang med hensyn til varighed, klassetrin og antal elever.
Helårseffekt	Effekt fra litteraturen omregnet (lineært) til effekt ved fuld implementering i et år.
Intensiv margin	Intensitet, med hvilken en indsats implementeres i løbet af skoleåret.
Konsekvens, konsekvensforbrug, konsekvensmål	Et udfald, til hvilken der kan knyttes økonomi, og som er relevant i statslig, regional eller kommunal budgetmæssig sammenhæng. Dækker forbrug af offentlige ydelser, services og aktiviteter.
Konsekvens-estimat	Estimatet på den betingede korrelation mellem link outcome og konsekvens.
Link outcome	Et udfald, der påvirkes af indsatsen og benyttes i litteraturen, men til hvilken der ikke kan knyttes økonomi. Bruges som bindeled mellem indsats og konsekvens.
Lokalt niveau	De parametre, der indgår i beregningen, måles på kommunalt niveau.
Outcome	Se konsekvens.
Persistens	Et mål for, hvor vedholdende effekten af en indsats kan forventes at være over tid. Bestemmes empirisk som korrelationen mellem link outcome i år $t = 1$ og konsekvens i år $t = 2, 3, 4$.
Pris	Enhedsprisen, der knytter sig til konsekvenserne. Benyttes til at knytte de estimerede indsatseffekter på forbrug af konsekvenser til økonomi.
Timing	Tidsmæssig skævhed mellem indsatsperiode og budgetperiode korrigeres med en korrektionsfaktor. Default i modellen er opstart af indsats efter sommerferien frem for ved budgetårets start, da indsatser på skoleområdet typisk følger skoleåret. Korrektionsfaktoren er 0,455.

Bilag 2 Vægte

Bilagstabel 2.1 Oversigt over vægte til brug i vægtningsmodul

Vægtalgoritme							
4 niveauer		3 niveauer		2 niveauer		Klassetrin	
Niveau	Vægt	Niveau	Vægt	Niveau	Vægt	Afstand	Vægt
0	0	0	0	0	0	0	1,00
1	0,50	1	0,50	1	0,5	1	0,80
2	0,75	2	0,82	2	1	2	0,67
3	0,90	3	1,00			3	0,57
4	1,00					4	0,50
						5	0,44
						6	0,40
						7	0,36
						8	0,33
						9	0,31

Bilag 3 Oversigt over klassificering af studier

Bilagstabel 3.1 Klassifikation af studier

Faktor		Kvalitet			Kontekst			Sammenlignelighed		
Subfaktor	Udgivelsesniveau	Empirisk metode	Datagrundlag	Geografisk område	Tidspunkt for indsats	Selekteret målgruppe	Indsats	Outcome	Tidshorisont	Klasse-trin
Klassificering	BFI 3	Lodtrækningsforsøg	Treatment-group større end 100 individer	Danmark	Fra 1990 og frem	Ikke høj grad af selektion	Samme indsats i model og studie	Samme outcome i model og studie	Outcome målt op til og med et år efter indsats	Afstand
	BFI 2	Kvasiek-sperimentelt	Treatment-group mindre end 100 individer	Nordiske lande	Før 1990	Høj grad af selektion	Indsats i studiet vurderes i nogen grad at overlape med modelindsatsen	Outcome i studiet vurderes i nogen grad at overlape med modeloutcome	Outcome målt 1-2 år efter indsats	
	BFI 1	Paneldata-model		Andre lande			Andet	Andet	Outcome målt senere	
		Andet								

VIVÉ