

# Evaluering av tiltak for å redusere sykefravær

Stein Atle Lie

Unifob helse, Christiesgate 13, 5015 Bergen, Norway

E-post: Stein.Lie@unifob.uib.no

## SAMMENDRAG

Evaluering av tiltak for å redusere sykefravær er nødvendig. Men hvorledes skal man måle effekten av et tiltak, hvilke undersøkelsesmetodikk og studiedesign skal man benytte og hvilke statistiske metoder skal anvendes? Ulike forsøksdesign gir ulik informasjon og kvalitet av informasjon, og dette bør settes i et perspektiv før man analyserer og tolker resultat fra tilgjengelige data. Denne artikkelen diskuterer ulike forsøksdesign delvis basert på kunnskapsbaserte metoder. Videre diskuteres hva som finnes av tilgjengelige data for sykefravær og hvilket perspektiv disse dataene kan settes i, i forhold til de ulike studiedesign. Avhengig av hvorledes et utfall (for eksempel tilbake til arbeid) er operasjonalisert er det ulike statistiske metoder som er naturlig for de tilgjengelige dataene og dette settes inn i perspektivet. Sist argumenteres det for å benytte betraktninger og metodikk fra utvidede forløpsanalyser i flertilstandsmodeller (multi-state modeller). Ved flertilstandsmodeller kan man videre utnytte en større del av dataene man har tilgjengelig i evaluering av et tiltak for å redusere sykefravær eller effekten av rehabiliteringstiltak.

## INNLEDNING

Stadig leser man i aviser og media om det høye sykefraværet i Norge. Et høyt sykefravær som på den ene siden problematiseres som et resultat av de gode sykelønnsordningene i Norge og mulighetene for at det utnyttes, mens det på den annen side er en viktig forsikring ved sykkelighet.

Det er satt i gang en lang rekke tiltak for å øke tilstedeværelse på jobben (nærvær), redusere fravær og lengden på fravær i bedrifter og arbeidslivet generelt og nye tiltak dukker jevnlig opp. Treningsopplegg i regi av jobben er et eksempel på tiltak som har et fokus på de som er i arbeid og som prøver å øke fysiske og psykiske prestasjoner og velvære for å bevare og beholde personer i arbeid<sup>1,2</sup>. For personer som er falt ut av arbeidslivet, for eksempel ved langvarig sykefravær, eksisterer det også en rekke tiltak med ambisjoner om å få personer tilbake til arbeidslivet. Kurs, omskolering, attføring, inkluderende arbeidsliv, raskere tilbake osv. er noen eksempler på slike tiltak. Noen av tiltakene, slik som inkluderende arbeidsliv, omfavner både de som er i arbeid og de som ikke er det.

Om de ulike tiltakene faktisk virker, for hvem og hvor mye er diskutert. Tveito og medforfattere viser at kun 10% av personer i arbeidslivet står for 82% av fraværet<sup>3</sup>. Dette kan brukes som et eksempel på at en god idé som trening på arbeidsplassen kan ha minimal effekt siden hovedmålgruppen, de med høyt fravær, ikke er til stede og ikke benytter seg av opplegget.

Årlig utbetales det store milliardbeløp i ulike ytelser fra NAV. Tall fra NAV viser at det for de tre første kvartalene i 2008 ble utbetalt 22 milliarder i sykepenger, mens de tilsvarende tallene for attføring/rehabilitering og uføretrygd var 14 og 44 milliarder kroner (<http://www.nav.no/98522.cms>). Ofte betraktes syke-

fravær i Norge på et makronivå ved å se på det totale sykefraværet i tverrsnitt ved eksempelvis hvert årskvartal (<http://www.nav.no/page?id=1073743235>). Dette er selvsagt nødvendig og viktig både for budsjettering og for å få et overordnet perspektiv på problemet. Men for evaluering av ulike tiltak som settes i verk er dette perspektivet langt fra ideelt. Det skal svært drastiske tiltak til for å gi en synlig endring i det generelle sykefraværet på kort sikt. Samtidig får man store fortolkningsproblemer på grunn av generelle konjunkturer og multiple faktorer som bidrar til det store bildet. Et eksempel på et drastisk tiltak med en synlig effekt på makronivå var innføring av karenisdag i Sverige i 1993<sup>4</sup>. Dette medførte at korttids sykefravær umiddelbart gikk ned, men etter noen år gikk sykefraværet opp igjen til det nivået det hadde før innføringen. Sykefraværet i Norge ligger på et svært høyt nivå sammenlignet med andre land, kun Sverige og Nederland ligger på et tilsvarende nivå<sup>4</sup>. Denne forskjellen forklares vanligvis ved de gunstige sosiale ordninger og den lave arbeidsledigheten som Norge har. Selv etter innføring av karenisdag kan man nok anse at sykelønnsordningen i Sverige var svært gunstig.

I tredje kvartal i 2008 var det 37,2% med oppgitt muskelskjelettlidelser som diagnose mens 13,9% hadde oppgitt lettere psykiatriske lidelser som årsak til sykefravær (<http://www.nav.no/155111.cms>). Hva som faktisk ligger bakenfor disse diagnosene er derimot ikke klart. Ofte vil det være mer enn en diagnose som er aktuell ved en sykemelding, hvis personen har sammensatte eller uklare problemer eller lidelser<sup>5,6</sup>. Samtidig er det en kjensgjerning at det må settes en diagnose for at personen skal få rettigheter til sykelønn. De heterogene forklaringsmekanismene gjør det komplisert å danne gode hovedprinsipp for å øke deltagelsen i arbeidslivet.

Det er nok ingen enkel løsning for å redusere sykefraværet og man vil neppe se en merkbar reduksjon av sykefraværet med det første. Men det er grunn til å håpe at personer som kommer eller er i en slik situasjon der de trenger det, får et tilbud som er dokumentert å ha en effekt og ikke et tilbud man kun tror eller håper har en virkning. Det er grunn til å tro at veloverveide og på forhånd gjennomtenkte studier på sikt kan føre oss et stykke på vei i dokumentasjon av effekten for ulike tiltak.

Formålet med denne artikkelen er flerdelt, først benyttes tanker fra kunnskapsbasert medisin (evidensbasert medisin, EBM) for å diskutere hva man faktisk kan regne med å finne ut basert på ulike typer studier og studiedesign. Dernest vil tilgjengelige datakilder drøftes opp mot de ulike studiedesign og hvilke statistiske metoder som faller naturlig. Til sist drøftes utvidede forløpsanalysemetoder for flertilstandsmodeller (multi-state modeller) og muligheten for å benytte dette for analyse av sykefraværdata.

## ULIKE TYPER STUDIER OG DATAGRUNNLAG

### Studie- og forsøksdesign

Concato og kolleger rangerer graden av kunnskap (evidens) for kliniske studier ut fra hvilket datagrunnlag studiene har (tabell 1)<sup>7</sup>. Innen kunnskapsbasert medisin danner en slik rangering grunnlaget for evaluering av kvaliteten for studier av behandling og helsetjenester. De randomiserte studiene (studier av klasse I) regnes som de eneste som gir konsistente (kausale) svar på effekter, der man vet at tiltaket (intervensjonen) kommer før utfallet. Innenfor Cochrane-samarbeidet ([www.cochrane.no](http://www.cochrane.no)) danner randomiserte kliniske forsøk ofte grunnlaget i de systematiske oversiktene. Denne typen rangering er ikke bare vanlig innen klinisk utprøving av medikamenter, men også innen en rekke andre fagområder der man ønsker et klart definert svar på effekten av en behandling (positiv eller negativ).

**Tabell 1.** Gradering/rangering av ulike typer studier.

Klassifisering	Beskrivelse
I	Kunnskap fra minst en godt designet kontrollert randomisert studie
II-1	Kunnskap fra godt designede forsøk uten randomisering
II-2	Kunnskap fra godt designede kohort eller kasus-kontroll studier
II-3	Kunnskap fra multiple tidsrekker, med eller uten intervensjon, samt dramatiske effekter i ukontrollerte forsøk.
III	Ekspertuttalelser basert på ekspertise, deskriptive studier eller kasusrapporter eller rapporter fra ekspertutvalg.

Norsk oversettelse av tabell 1 i Concato et al.<sup>7</sup> for klassifisering av evidens (kunnskap) fra ulike typer studier.

I den nå klassiske studien hevder Concato og kolleger<sup>7</sup> videre at gode observasjonsstudier ofte gir de samme resultatene som det man finner i randomiserte kontrollerte forsøk, selv om det finnes eksempler på at dette ikke alltid er tilfellet. Ved kontrollerte og randomiserte forsøk kan man og få motsatt resultat av det som ble konkludert fra observasjonsstudier<sup>8,9</sup>. Det er verdt å merke seg her at effekter fra randomiserte studier ofte vil omtales som kausale, mens effekter fra andre typer studier vanligvis omtales som assosiasjoner.

Man kan selvsagt argumentere både for og imot rangeringen i tabell 1 og man kan si at den gir en for grov inndeling eller at noe mangler. Likevel gir tabellen et utgangspunkt for å diskutere ulike typer studier og muligheten slike studier har til å fange opp og måle reelle endringer – i for eksempel sykefravær.

En rekke av de overordnede og politisk vedtatte tiltak som settes inn for å redusere sykefravær kan nok kategoriseres som studier av type III, siden de ofte har en svak empirisk begrunnelse. Selv om det ikke er tenkt som studier vil man likevel svært ofte argumentere for effekter når slike tiltak settes i verk, eller at tiltaket evalueres basert på en subjektiv oppfatning hos en liten gruppe personer. Evalueringen av overordnede tiltak baserer seg ofte på om at man kan måle effekter innenfor kategori II-3. Dette er ikke ment å være et generelt motargument for overordnede tiltak eller måten det evalueres på, men for å synliggjøre hvor det kan passe inn i en slik stringent oversikt. Generelt er det nettopp sykefravær på et makronivå som er markøren for befolkningens bruk av denne ytelsen. For at et tiltak skal ha en målbar effekt på det totale sykefraværet skal effekten være stor, samtidig som man risikerer at effekten maskeres (både i negativ og positiv retning) av andre endringer, for eksempel endringer i regelverk eller i generelle konjunkturer.

Et godt eksempel på et overordnet tiltak med en klar effekt på reduksjon av sykefraværet var innføringen av karensdag i Sverige i 1993. I hvert fall er man rimelig sikker på at det var forklaringen av nedgangen i sykefravær, siden nedgangen fulgte i tiden etter innføringen av tiltaket. Etter flere år med nedgang i sykefravær i Sverige steg igjen sykefraværet tilbake til det samme nivået som før innføringen. Grunnene til dette kan være flere, både økonomiske og medisinske, for eksempel at ytelsene for sykefravær fortsatt kan regnes som svært gode eller at befolkningens grad av sykkelighet over tid stabiliserer seg på et gitt nivå uavhengig av de økonomiske insentivene. Uansett viser dette at graden av kunnskap basert på effektene man observerte ikke når høyere enn nivå II-3 i Concatos oversikt i tabell 1.

### Datagrunnlag

Datakildene for sykefravær, stønader og trygdeytelser er tilgjengelig på individbasert nivå i Norge, for eksempel i de administrative databasene som NAV forvalter eller datasamlingen FD-trygd (forløpsdata trygd, <http://www.ssb.no/emner/03/fd-trygd/>) ved Statistisk sentralbyrå. Individbaserte data åpner for flere meto-

diske muligheter og utfordringer. Man får for eksempel muligheten til å følge enkeltpersoners vandring mellom ulike ytelser. Videre får man også mulighet til å koble denne typen informasjon til andre kilder med individbasert informasjon, som det har vært gjort i en rekke studier<sup>10-17</sup>. Norge er i en unik situasjon internasjonalt for å gjøre studier med kobling av data fra ulike kilder på grunn av det unike personnummer i Norge og våre individbaserte befolkningsdatabaser.

Hvis vi nå ser tilbake til oversikten fra Concato ser vi at slike register data i seg selv ikke øker klassifiseringen for graden av evidens, men potensialet for å gjøre det er til stede. Hvis man for eksempel kobler individbaserte data om ytelser til opplysninger om hvilke tiltak personer har vært gjennom kan man muligens argumentere for å være på nivå II-2 innen evidens. Hvis informasjonen om tiltak er svært spesifikk og man dessuten har tilstrekkelig tilleggsinformasjon kan man muligens si at man har evidens av klassifisering II-1 for evalueringen av det tiltaket.

Fortsatt er det slik at det bare er gode randomiserte kontrollerte studier som gir presis kunnskap om den kausale effekten av et tiltak. Med dette menes at for alle andre typer studier må man ta forbehold om at skjevheter (bias), seleksjoner eller annet kan påvirke effekten av tiltaket man studerer. Dette gjelder selv om man har foretatt justeringer i de statistiske analysene eller på annen måte. For randomiserte kontrollerte forsøk kan man selvsagt med fordel følge personer i etterkant ved å innhente data fra de individbaserte databasene som finnes ved NAV og FD-trygd.

Senere i artikkelen konkretiseres problemene og metodene ved eksempler på studier og datatype, fra en randomisert studie med data fra sykehuset Innlandet<sup>18,19</sup> og en observasjonsstudie med data fra Attføringsinstituttet i Rauland.

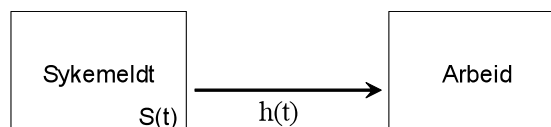
### Statistiske metoder

Det er ikke intensjonen å gi en utfyllende og generell oversikt over statistiske metoder her, men å gi noen eksempler på hvilke metoder som benyttes og videre komme med noen anbefalinger. De statistiske metodene man benytter vil selvsagt avhenge av de data man har tilgjengelig, eller har samlet inn, og hvorledes disse er organisert og operasjonalisert.

Innen helse og trygdeøkonomi ser man ofte på hvorledes tverrsnittet av sykefravær, og eventuelt andre ytelser og konjunkturer, endrer seg over tid. Til dette benyttes det naturlig nok modeller for tidsrekker enten i enkle varianter eller som panel<sup>20</sup>.

For epidemiologiske studier av sykefravær benyttes de fleste standard statistiske metoder og modeller<sup>21,22</sup>. For eksempel benytter man ved opptelling av antall reduserte sykefraværsdager lineære metoder og modeller (t-test, ANOVA, GLM, osv.) eller Poisson-modeller, mens når man ser på et gitt tidspunkt og undersøker hva som er status da (for eksempel om en person er i arbeid ved 1 år) benytter krysstabeller (og kji-kvadrat-tester) eller logistisk regresjon<sup>22</sup>.

For de fleste ytelser sier det seg selv at tid er en vesentlig faktor, derfor benyttes i tiltagende grad ulike former for forløpsanalyse til å studere effekter på det å returnere til jobb. Metoder som Kaplan-Meier estimat for å beregne overlevelsessannsynligheter er utbredt. For beregning av overlevelsesfunksjon antar man at problemstillingen man står overfor kan skisseres som i figur 1. I denne situasjonen har man for eksempel en gruppe individer som er sykemeldt ved observasjonens start og som man ønsker å følge inntil de returnerer til arbeid.



**Figur 1.** Skjematisk framstilling av en enkel forløpsanalyse for skifte fra sykefravær til arbeid.

Sannsynligheten ( $S(t)$ ) for fortsatt å være sykemeldt er definert til å være 1 (100%) ved tid 0 og avtar deretter monotont, siden ingen kan returnere til starttilstanden. Cox proporsjonale hazarder regresjon (og også andre hazard-regresjonsmodeller) benyttes ofte for å se på risikofaktorer for å skifte fra starttilstanden (sykemeldt) til stopptilstanden (arbeid). Cox-regresjon gir risikoestimat som er ratioen av hazardene ( $h(t)$ ) for ulike kategorier av variable (under forutsetning av at denne sammenhengen er proporsjonal over tid). Kaplan-Meier beregning av sannsynligheter og Cox-regresjon er svært utbredt og finnes i alle standard statistikkprogrammer. I utgangspunktet vil man i enkle forløpsanalyse-modeller kun fokusere på første hendelse i en forløp, for eksempel første gang personen returnerer til arbeid. Utfallet uføretrygd kan i så henseende være et hensiktsmessig endepunkt, siden det ofte kan oppfattes som en absorberende tilstand, som personer i liten grad forlater<sup>11,15-17</sup>. Hvis en person skifter gjentatte ganger mellom arbeid og sykemeldt, som skissert i figur 2, er det altså oftest kun den første gangen personen skifter fra sykemeldt til arbeid som inkluderes i analysene.

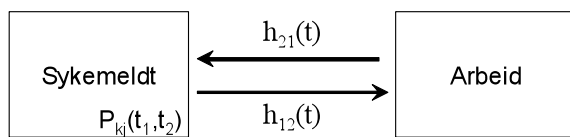
For å håndtere gjentatte observasjoner for et individ eller korrelerte data innen forløpsanalyser finnes det ulike muligheter<sup>23-25</sup>.

### Flertilstandsmodeller (multi-state modeller)

For data innen oppfølging av personer med sykefravær og innen andre ytelser vil man generelt være interessert i hvilket forløp personene faktisk har. Den enkleste situasjonen kan være at personer skifter gjentatte ganger mellom det å være sykemeldt og det å være i arbeid, slik som gjengitt i figur 2.

Forskjellen på situasjonen i figur 2 og den enklere i figur 1 er at her tillater vi at personer gjentatte ganger kan bevege seg mellom de to tilstandene. For denne utvidete situasjonen har vi 2 hazardfunksjoner, den ene

indikerer intensiteten for å bevege seg fra tilstand 1 til tilstand 2 ( $h_{12}(t)$ ), mens den andre er intensiteten for å skifte den andre veien ( $h_{21}(t)$ ). Videre vil vi nå få 4 ulike overgangssannsynligheter ( $P_{kj}$ ) som er sannsynlighetene for å være i en gitt tilstand (1 eller 2) ved et tidspunkt ( $t_2$ ) avhengig av hvilken tilstand (1 eller 2) personen var i ved et tidligere tidspunkt ( $t_1$ ), uavhengig av antall skifter mellom de to tidspunktene. En person som er sykemeldt ved tidspunkt 1 kan ved tidspunkt 2 være sykemeldt eller i arbeid og kan i tiden mellom de to tidspunktene ha hatt gjentatte sykemeldings- og arbeidsperioder. Overgangssannsynlighetene setter man vanligvis opp i en overgangssannsynlighetsmatrise ( $\mathbf{P}$ ), slik som i formel 1 for den relativt enkle situasjonen med kun to mulige tilstander.



**Figur 2.** Skjematisk framstilling av en utvidet forløpsanalyse for gjentatte skift mellom sykefravær og arbeid.

$$\mathbf{P}(t_1, t_2) = \begin{pmatrix} P_{11}(t_1, t_2) & P_{12}(t_1, t_2) \\ P_{21}(t_1, t_2) & P_{22}(t_1, t_2) \end{pmatrix}$$

**Formel 1.** Matrise for overgangssannsynligheter for alle mulige sannsynligheter i modellen i figur 2.

Beregningen av matrisen  $\mathbf{P}$  gjøres ofte ved produktintegral (Aalen-Johansen estimatoren<sup>26</sup>) og baserer seg på overgangsintensitetene (hazardene) for å skifte mellom de ulike tilstandene (ofte beregnet ved Nelson-Aalen estimatoren<sup>27,28</sup>). For noen spesialtilfeller av flertilstandsmodeller kan beregningen av overgangssannsynlighetene forenkles eller gjøres direkte. For eksempel vil modellen i figur 1 være det enkleste av slike spesialtilfeller, der Kaplan-Meier estimatoren blir et spesialtilfelle av en overgangssannsynlighet. Grunnprinsippet innen flertilstandsmodeller for beregning av overgangssannsynligheter, som er skissert her, kan selvsagt utvides til mer komplekse modeller og datastrukturer med flere tilstander. Videre blir også mulighetene for å gjøre regresjonsmodeller utvidet.

Det er viktig å presisere at flertilstandsmodeller ikke i seg selv løser noen problemer knyttet til diskusjonen av klassifisering av studier, men at det kan vise seg å være et nyttig verktøy innen studier både innen klassifisering I, II-1 og II-2.

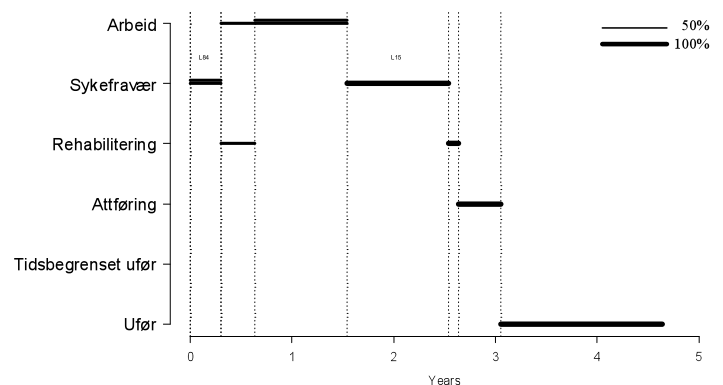
## RESULTAT FRA EKSEMPLER

I en nylig publisert artikkel benyttet Lie og medforfattere en 3-state modell (RTV-ytelse, arbeid og uføretrygd), med 3-års oppfølging, som utfall for en

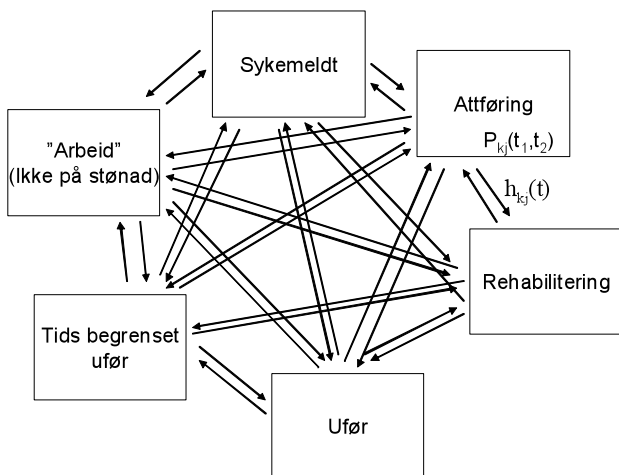
randomisert kontrollert studie<sup>29</sup> med data fra sykehuset Innlandet. Artikkelen baserer seg på 456 personer, som hadde vært langtidssykemeldt (mer enn 8 uker) på grunn av korsryggsmerter. Pasientene ble randomisert til en kortvarig intervensjon (brief intervention) på bare 4 timer, som blant annet inkluderte en grundig medisinsk undersøkelse og informasjon. Analysene fant at det underveis i 3-års perioden var en forskjell i sannsynligheten for å være i arbeid (og også for å være på en ytelse fra NAV), mens det ved 3 år ikke var en slik forskjell<sup>29</sup>. Den observerte forskjellen i sannsynligheter underveis i 3-års perioden kan også presenteres som et visst antall sparte (eller tapte) dager per person. I denne studien fant vi at antall dager ekstra i arbeid for personer i intervensjonsgruppen var 48,9 dager (95% CI: 22,6–77,6 dager).

Fra Attføringsinstituttet i Rauland (AiR) har vi fulgt en årsproduksjon (2001) i 5 år etter de forlot rehabiliteringsopplegget uten en kontrollgruppe. Totalt var det 581 personer som hadde et rehabiliteringsopphold ved Rauland i 2001 og de fleste av disse var på grunn av muskel- og skjelettplager og lettere psykiatri. De fleste personene ( $n=364$ ) stod på sykepenger da de forlot AiR. Oppfølgingen av personene ble gjort ved å koble data på individnivå med opplysninger om ulike typer ytelser fra NAV. For denne problemstillingen satte vi opp en modell med 6 ulike tilstander; sykemeldt, attføring, rehabilitering, tidsbegrenset ufør, ufør og i arbeid. Definisjonen av arbeid er her at de ikke er registrert med en ytelse fra NAV. Denne definisjonen kan selvsagt diskuteres da personer kan være alt annet enn i arbeid, men for alle praktiske formål og for illustrasjonens del synes vi denne definisjonen er adekvat. For illustrasjon har vi i figur 3 data fra en tilfeldig person og dennes vandring mellom ulike ytelser og arbeid.

En skisse av en flertilstandsmodell for denne problemstillingen er satt opp i figur 4, som viser at det er et stort antall hazardfunksjoner (piler/intensiteter) som er definert og at det dermed også er svært mange overgangssannsynligheter for de ulike mulige utfall i modellen.



**Figur 3.** Eksempel på vandring mellom ulike ytelser for en person, som til sist blir uføretrygdet.



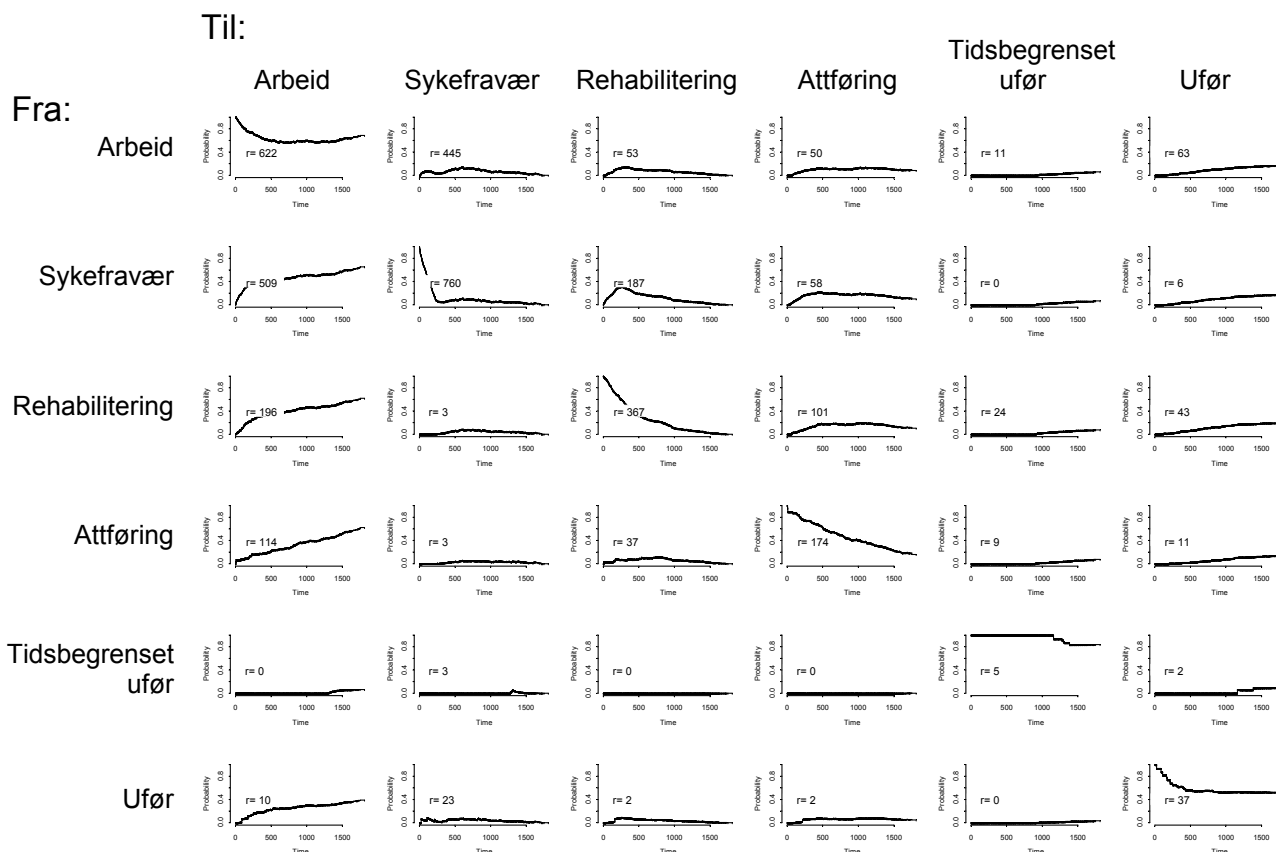
Figur 4. Skisse av en flertilstandsmodell for vandring mellom ulike ytelser og arbeid.

Resultatet fra denne analysen kan presenteres som en 6x6 matrise med 36 ulike overgangssannsynlighetsfunksjoner som vist i figur 5. I denne figuren kan man for eksempel lese av hva sannsynligheten er for å være i hver enkelt av de 6 mulige tilstandene etter avreise fra Rauland, gitt at man vet hvilken status personene hadde ved avreise. For eksempel kan man se sannsynligheten for å være i arbeid i 5-års perioden etter avreise, hvis man var sykemeldt ved avreise fra AiR.

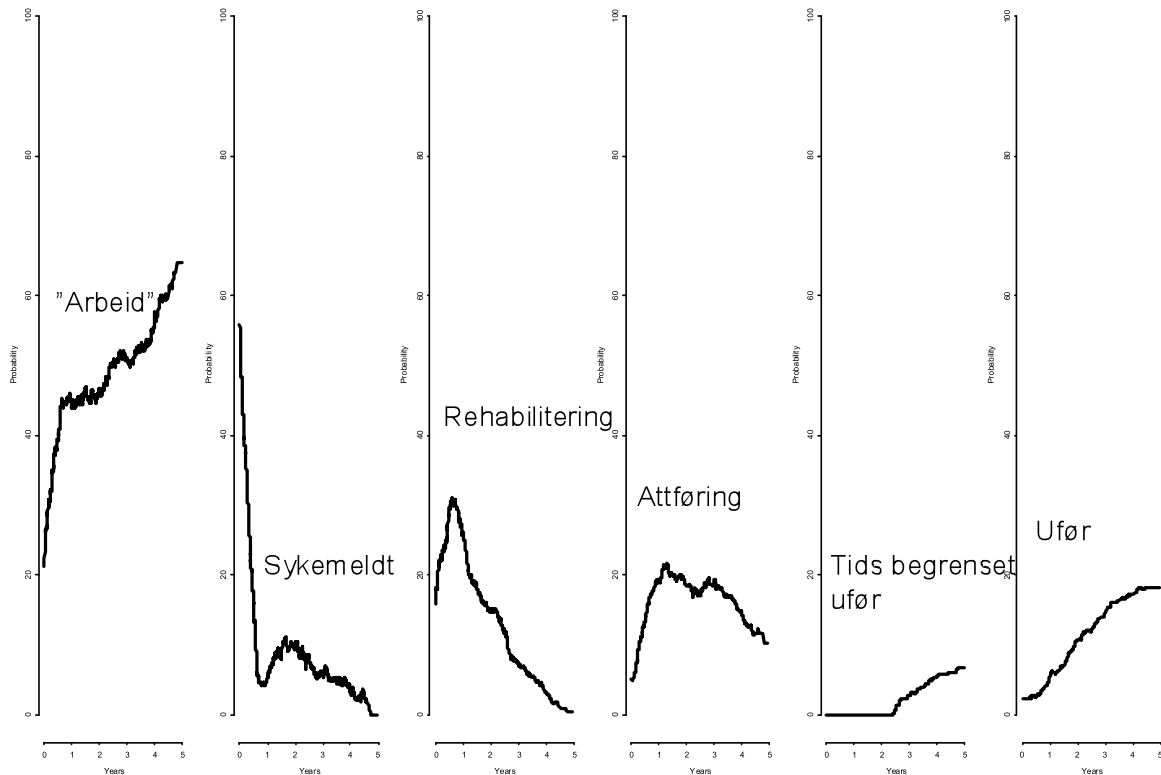
En slik oversikt kan bli stor og uoversiktlig. Videre

er det ikke sikkert at det er like interessant å få vite sannsynligheten for alle mulige kombinasjoner av utfall. Et alternativ til å vise disse simultane overgangssannsynlighetene er å beregne vektete sannsynligheter, som i figur 6 (her vektet med frekvensfordelingen for de ulike tilstandene ved tid 0), for å gi en total oversikt over det forventede forløpet til en tilfeldig valgt person etter avreise, uavhengig av hvilken tilstand personen var i ved avreise.

Selv om 60% tilbake i arbeid ved 3 år sikkert er et høyt tall og et bra resultat vet man likevel ikke om dette faktisk er bedre enn andre tilbud disse personene kunne fått. For å avgjøre om det faktisk er en effekt av et opphold ved AiR måtte man hatt en kontrollgruppe og aller helst burde det vært en tilfeldig (randomisert) fordeling mellom det å få et opphold ved AiR og alternativet. Om det er praktisk gjennomførbart å randomisere personer mellom det å dra til Rauland og noe annet er også et spørsmål, som kompliserer det å faktisk gjennomføre en slik studie. Det å sammenligne personer som har vært ved AiR med en ikke randomisert kontroll gruppe (for eksempel en annen klinikk) vil ikke nødvendigvis si noe om effekten av det å ha vært på AiR, men like mye hvem som selekteres (velger) å dra til AiR. Studier av denne typen kan likevel ha en høy relevans når det gjelder å se på ulike risikofaktorer. Å identifisere variable som indikerer om personen har effekt av et slikt opphold eller ikke kan være av høy verdi. Dette er et arbeid som pågår.



Figur 5. Sannsynlighetsfunksjoner for kombinasjoner (simultane) av alle mulige utfall etter avreise fra AiR.



Figur 6. Vektete (marginale) sannsynlighetsfunksjoner for ulike utfall etter avreise fra AiR.

## DISKUSJON

Denne artikkelen har diskutert og problematisert ulike sider av evaluering av tiltak for sykefravær og rehabilitering. Hovedintensjonen med artikkelen var å argumentere for at evaluering av tiltak er nødvendig og at en slik evaluering, og metodikken for denne, må være gjennomtenkt før tiltaket settes i verk i bred skala. Flere har påpekt at det er et kritisk behov for mer og bedre forskning på sykefravær og for hva som har effekt når det gjelder å få personer til å returnere til arbeid<sup>30,31</sup>. Videre er et svært viktig argument at når man skal evaluere tiltak for å få personer med sykefravær eller i et rehabiliteringsløp tilbake til jobb så må effektmålet når man gjør evalueringer være det at de faktisk er tilbake i jobb. At personer subjektivt er fornøyd med et opplegg for å returnere til arbeid betyr definitivt ikke at de øker sjansen sin for faktisk å komme tilbake til jobb, det motsatte kan faktisk være tilfelle. Innen medisin er et av de mest berømte eksempler på at surrogatendepunkt kan være direkte misvisende CAST-studien som viste at selv om et medikament for å redusere hjertearytmi virket helt etter intensjonen gav det en formidabel økning av dødelighet i disfavør av medikamentet<sup>32</sup>. Medikamentet var på tidspunktet for publikasjonen svært utbredt i USA, men ble da raskt trukket fra markedet.

”Tilbake til jobb” kan virke som et veldefinert og klart utfallsmål, men det er flere ulike måter å måle resultatet basert på dette utfallet. Hensing og medforfat-

tere diskuterer ulike mål knyttet til i arbeid som utfall<sup>22</sup> og selv om vi benytter oss av flertilstandsmodeller kan defineringen av resultatmålet variere. Det kan være vanskelig å måle hva det vil si å være i arbeid. I en artikkel av Øyeflaten og medforfattere i denne utgaven av *Norsk Epidemiologi*<sup>33</sup> diskuteres det om det er samsvar mellom egenoppgett status (fra spørreskjema) og det som er registrert i databasene ved NAV. Det var en relativt stor grad av overensstemmelse mellom de to ulike datakildene, men var det noen påfallende skjevheter. Personer kan selv oppfatte at de er i arbeid (for eksempel ved aktiv sykemelding eller attføring), mens NAV ikke vil definere dette som i arbeid.

Videre belyser denne artikkelen viktigheten av å ha en referanse for å måle hva effekten av et tiltak er. Concato et al. argumenterer for at hvis man i ukontrollerte studier gjør gode og veloverveide justeringer kan man oppnå en fortolkning av resultatene som ligner det man ville hatt i randomiserte studier<sup>7</sup>. Men det er fortsatt så mange uklare og uforklarte mekanismer innen årsaker til sykefravær, som man ikke kan justere for ved statistiske analyser, at man skal være svært varsom med å trekke konklusjoner. Et tiltak som iverksettes nasjonalt kan således i prinsippet svært vanskelig evalueres, fordi man normalt ikke har en relevant sammenligning (kontrollgruppe) og de personene som faktisk gjennomfører et slikt tiltak kan representere en seleksjon (skjevt utvalg). Personene som gjennomfører tiltaket trenger derfor ikke være sammenlignbare med de som ikke gjennomfører et

slikt tiltak og sammenligning av gruppene vil ikke være en evaluering av tiltaket, men av seleksjonen. Som en generell regel kan ukontrollerte studier (både observasjonsstudier og tidsserier) bare avsløre svært store effekter av et tiltak, som man generelt ikke vil ha innen dette fagfeltet. Det eneste tiltak man kanskje kan vente vil gi markante utslag i slike studier er sannsynligvis radikale endringer i stønadsordningene, for eksempel ved innføring av karensdager.

Det kan videre være svært nyttig for forskere, men også for personer innen forvaltning, å ha en viss forståelse for rangering av ulike typer studier, hva som kan kalles assosiasjoner og hva som kan kalles kausale sammenhenger. At kun randomiserte studier gir sanne kausale sammenhenger må sees i lys av hva som kreves for å gjennomføre slike studier opp mot studier som kan benytte det som allerede finnes av tilgjengelige data. Randomiserte forsøk er dyre, tidkrevende og de gir som regel kun svar på et tiltak om gangen. Data blant annet fra helseundersøkelser eller helseregistre koblet mot data fra NAV/FD-trygd eller data fra NAV/FD-trygd alene, kan sammen med adekvate statistiske metoder alternativt gi verdifull informasjon. Videre kan data fra disse kildene med fordel analyseres med metodikk fra flertilstandsmodeller. Uansett må man huske å ta forbehold om tolkning av resultatene i ikke randomiserte studier, gjerne sett i lys av en rangering av studier som gjengitt her.

Selv om det ofte kan være fornuftig å benytte flertilstandsmodeller for å utnytte de dataene man faktisk har tilgjengelig (for eksempel fra NAV eller FD-trygd) kan enklere modeller generelt gi det samme resultatet. For eksempel vil en opptelling av andel på ulike ytelser ved ulike tidspunkt (for eksempel ved 3, 6 og 12 måneder) i prinsippet gi de samme estimatene som

man ville finne basert på beregning av overgangssannsynligheter. Videre kan for eksempel logistiske regresjonsmodeller for gitt tidspunkt med fordel benyttes. Men selv om ikke det er lett tilgjengelig programvare for flertilstandsmodeller kan det være fornuftig å benytte tankesett fra dette, som skissert i figur 1, 2 og 4. Videre er en utvilsom fordel med forløpsanalyser (og da også flertilstandsmodeller) at man tar hensyn til at personer kan ha ulike oppfølgingstid (sensurering), eller være stykkevis observert (intervallsensurering), slik det vil være når man observerer ulike ytelser samtidig.

Det finnes nok ingen raske løsninger å redusere det høye sykefraværet i Norge. At flere ulike typer tiltak virker er meget mulig. Både holdningskampanjer, økonomiske incentiv og fokus på fysisk og psykisk helse er relevant for å holde personer i arbeidslivet. Men, for personer som er i ferd med å falle ut, eller har falt ut av arbeidslivet er det svært viktig å ha kunnskap om at tiltakene man iverksetter har en effekt, noe annet vil være uetisk.

For å få kunnskap om hva som er virkningsfulle tiltak, og hvor stor effekten av et tiltak er, må det i prinsippet gjennomføres randomiserte kontrollerte forsøk. Men, eksisterende datakilder i Norge gjør at man med fordel og med betydelig mindre innsats og ressurser kan gjøre andre typer studier, for eksempel observasjonsstudier der man følger store grupper av enkelt-individ med statistiske metoder for flertilstandsmodeller, eller tidsserieanalyser for aggregerte data i panel. Slike studier har og bør ha en sentral plass i evalueringsforskning og vil, hvis effektene av tiltak er tilstrekkelig store, kunne avdekke dette men med forbehold.

## REFERANSER

1. Tveito TH, Eriksen HR. Integrated health programme: a workplace randomized controlled trial. *J Adv Nur*, **65** (1): 110-9, 2009.
2. Eriksen HR, Ihlebaek C, Mikkelsen A, Grønningsaeter H, Sandal GM, Ursin H. Improving subjective health at the worksite: a randomized controlled trial of stress management training, physical exercise and an integrated health programme. *Occup Med*, **52** (7): 383-91, 2002.
3. Tveito TH, Halvosen A, Lauvålien JV, Eriksen HR. Room for everyone in working life? 10% of the employees – 82% of the sickness leave. *Norsk Epidemiologi*, **12** (1): 63-68, 2002.
4. Nyman K, Bergendorff S, Palmer E. Den svenska sjukan. Sjukfrånvaro i 8 länder. DS 2002; 49. Stockholm: Regeringskansliet, 2002.
5. Eriksen HR, Ihlebaek C. Subjective health complaints. *Scand J Psychol*, **43** (2): 101-3, 2002.
6. Ihlebaek C, Eriksen HR. Occupational and social variation in subjective health complaints. *Occup Med*, **53** (4): 270-8, 2003.
7. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med*, **342** (25): 1887-92, 2000.
8. Rossouw JE, et al. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results From the Women's Health Initiative randomized controlled trial. *JAMA*, **288** (3): 321-33, 2002.
9. Chlebowski RT, et al. Breast cancer after use of estrogen plus progestin in postmenopausal women. *N Engl J Med*, **360** (6): 573-87, 2009.

10. Gravseth HM, Bjerkedal T, Irgens LM, Aalen OO, Selmer R, Kristensen P. Influence of physical, mental and intellectual development on disability in young Norwegian men. *Eur J Public Health*, **18** (6): 650-5, 2008.
11. Gravseth HM, Bjerkedal T, Irgens LM, Aalen OO, Selmer R, Kristensen P. Life course determinants for early disability pension: a follow-up of Norwegian men and women born 1967-1976. *Eur J Epidemiol*, **22** (8): 533-43, 2007.
12. Gjesdal S, Maeland JG, Svedberg P, Hagberg J, Alexanderson K. Role of diagnoses and socioeconomic status in mortality among disability pensioners in Norway – a population-based cohort study. *Scand J Work Environ Health*, **34** (6): 479-82, 2008.
13. Gjesdal S, Svedberg P, Hagberg J, Alexanderson K. Mortality among disability pensioners in Norway and Sweden 1990-96: comparative prospective cohort study. *Scand J Public Health*, **37** (2): 168-75, 2009.
14. Kristensen P, Bjerkedal T, Irgens LM. Early life determinants of musculoskeletal sickness absence in a cohort of Norwegians born in 1967-1976. *Soc Sci Med*, **64** (3): 646-55, 2007.
15. Mykletun A, Heradstveit O, Eriksen K, Glozier N, Overland S, Maeland JG, Wilhelmsen I. Health anxiety and disability pension award: The HUSK Study. *Psychosom Med*, **71** (3): 353-60, 2009.
16. Sivertsen B, Overland S, Pallesen S, Bjorvatn B, Nordhus IH, Maeland JG, Mykletun A. Insomnia and long sleep duration are risk factors for later work disability. The Hordaland Health Study. *J Sleep Res*, **18** (1): 122-8, 2009.
17. Overland S, Glozier N, Henderson M, Maeland JG, Hotopf M, Mykletun A. Health status before, during and after disability pension award: the Hordaland Health Study (HUSK). *Occup Environ Med*, **65** (11): 769-73, 2008.
18. Hagen EM, Eriksen HR, Ursin H. Does early intervention with a light mobilization program reduce long-term sick leave for low back pain? *Spine*, **25** (15): 1973-6, 2000.
19. Hagen EM, Grasdahl A, Eriksen HR. Does early intervention with a light mobilization program reduce long-term sick leave for low back pain: a 3-year follow-up study. *Spine*, **28** (20): 2309-15; discussion 2316, 2003.
20. Askildsen JE, Bratberg E, Nilsen OA. Unemployment, labor force composition and sickness absence: a panel data study. *Health Econ*, **14** (11): 1087-101, 2005.
21. Navarro A, Reis RJ, Martin M. Some alternatives in the statistical analysis of sickness absence. *Am J Ind Med*, **52** (10): 811-6, 2009.
22. Hensing G, Alexanderson K, Allebeck P, Bjurulf P. How to measure sickness absence? Literature review and suggestion of five basic measures. *Scand J Soc Med*, **26** (2): 133-44, 1998.
23. Aalen OO. Effects of frailty in survival analysis. *Stat Methods Med Res*, **3** (3): 227-43, 1994.
24. Therneau TM, Grambsch PM. Modeling survival data: extending the Cox model. New York, Springer, 2000.
25. Aalen OO, Bjertness E, Soonju T. Analysis of dependent survival data applied to lifetimes of amalgam fillings. *Stat Med*, **14** (16): 1819-29, 1995.
26. Aalen OO, Johansen S. An empirical transition matrix for nonhomogeneous Markov chains based on censored observations. *Scand J Stat*, **5**: 141-50, 1978.
27. Andersen PK, Keiding N. Multi-state models for event history analysis. *Stat Methods Med Res*, **11** (2): 91-115, 2002.
28. Andersen PK, Borgan B, Gill RD, Keiding N. Statistical Models Based on Counting Processes. New York, Springer-Verlag, 1993.
29. Lie SA, Eriksen HR, Ursin H, Hagen EM. A multi-state model for sick-leave data applied to a randomized control trial study of low back pain. *Scand J Public Health*, **36** (3): 279-83, 2008.
30. Alexanderson K, Hensing G. More and better research needed on sickness absence. *Scand J Public Health*, **32** (5): 321-3, 2004.
31. Pransky G, Gatchel R, Linton SJ, Loisel P. Improving return to work research. *J Occup Rehab*, **15** (4): 453-7, 2005.
32. Echt DS, et al. Mortality and morbidity in patients receiving encainide, flecainide, or placebo. The Cardiac Arrhythmia Suppression Trial. *N Engl J Med*, **324** (12): 781-8, 1991.
33. Øyeflaten I, Lie SA, Ihlebæk C, Brage S, Eriksen HR. Are self reported data on sick leave reliable? – A comparison between self reported and register data. *Norsk Epidemiologi*, **19** (2), 161-7, 2009.