

Rapport

# **Sykmelding, gradert sykmelding og forløpet tilbake til jobb**

**Gjentatte overganger mellom full sykmelding, gradert  
sykmelding og jobb for 283 296 personer fra 2002 til 2010.**

**Utarbeidet av**  
Stein Atle Lie, Uni helse

## Forord

Denne rapporten baserer seg på diskusjoner og analyser gjennomført i løpet av et kurs/undervisningsopplegg Uni helse har hatt med Arbeids og velferdsdirektoratet (NAV). Deltagere på kurset var fra seksjonene resultatoppfølging og analyse og statistikk og utredning. Undervisningsopplegget ble gjennomført i løpet av høsten 2009 og våren 2010.

Undervisningsopplegget hadde primært fokus på kunnskapsheving av personell tilknyttet de to seksjonene innen forløpsanalyser og bruk av nyere statistiske metoder innenfor forløpsanalyser (flertilstandsmodeller). Undervisningsopplegget bestod av to hoveddeler. Den første innledende delen fokuserte på innføring i forskningsmetodikk, beskrivelse av pasienthistorier, diagnostisering av sammensatte plager samt årsaker til sykefravær og subjektive helseplager. Den vesentligste delen av undervisningsopplegget bestod av introduksjon i forløpsanalyser, generelle formler, begrep samt relasjonen mellom dem innenfor enkle forløpsmodeller. Videre ble det fokusert på utvidelse av enkle forløpsanalyser til flertilstandsmodeller og applisering av flertilstandsmodeller på data fra NAV med eksempler. Aktuelle problemstillinger på data fra NAV ble diskutert opp mot organisering av dataene og telleprosess representasjon av disse. Analysene som foreligger i denne rapporten baserer seg på dette. Undervisningsopplegget ble ledet av Stein Atle Lie ved Uni helse, datauttrekkene i rapporten er gjort av Ola Thune ved seksjon for statistikk og utredning.

Problemstillingene og de analytiske tilnærmingene i denne rapporten er basert på diskusjonene i gruppen som deltok på undervisningsopplegget. Det er tatt sikte på at dette skulle bidra til kunnskapshevingen og en bredere forståelse av forløpsanalyser, samt at tilsvarende analysemetoder skal kunne diskuteres og benyttes også på andre områder, for eksempel i analyser av arbeidsavklaringspenger. Det ble også diskutert muligheter for å definere kvalitets-indikatorer basert på forløpsdata og problemstillinger i denne rapporten. NAV har en rekke problemstillinger der forløpsanalyser er en viktig tilnærming og en kompetanseheving på dette området anses derfor som viktig. En spesiell takk til seksjonsleder Inger Spangen.

Prosjektet er finansiert av NAV gjennom programmet FARVE - forsøksmidler arbeid og velferd.

## Innholdsfortegnelse

<a href="#">Innholdsfortegnelse .....</a>	<a href="#">3</a>
<a href="#">Sammendrag.....</a>	<a href="#">4</a>
<a href="#">1. Innledning.....</a>	<a href="#">6</a>
<a href="#">2. Datautvalg.....</a>	<a href="#">7</a>
<a href="#">2.1. Flertilstandsmodell .....</a>	<a href="#">9</a>
<a href="#">3. Statistisk metode.....</a>	<a href="#">11</a>
<a href="#">3.1. Overgangsintensiteter.....</a>	<a href="#">11</a>
<a href="#">3.2. Overgangssannsynligheter.....</a>	<a href="#">12</a>
<a href="#">3.3. Tilstandssannsynligheter.....</a>	<a href="#">13</a>
<a href="#">3.4. Tilstandsforventning.....</a>	<a href="#">14</a>
<a href="#">3.5. Regresjonsmodeller.....</a>	<a href="#">14</a>
<a href="#">4. Resultater.....</a>	<a href="#">15</a>
<a href="#">4.1. Totale resultater.....</a>	<a href="#">15</a>
<a href="#">4.2. Resultat for kjønn.....</a>	<a href="#">20</a>
<a href="#">4.3. Resultat for alder.....</a>	<a href="#">23</a>
<a href="#">4.4. Resultat for kalenderår.....</a>	<a href="#">26</a>
<a href="#">4.5. Oppsummering.....</a>	<a href="#">29</a>
<a href="#">5. Diskusjon.....</a>	<a href="#">30</a>
<a href="#">6. Oppsummering/Konklusjon.....</a>	<a href="#">33</a>
<a href="#">Referanser.....</a>	<a href="#">34</a>

## Sammendrag

I diskusjonen om sykefraværet og hvordan det kan reduseres er ”gradert sykmelding” (mindre enn 100 % sykmeldt) forslått som et ønsket tiltak. Det antas at dette vil redusere fraværet og at gradert fravær er å foretrekke framfor en full sykmelding. Hensikten er å nyttiggjøre at personen fortsatt har en viss arbeidsevne samt at personen holdes nærmest mulig arbeidslivet i håp om at dette kan lede til at personen senere kommer fullt tilbake til jobb. Det er lite eller ingen empirisk begrunnelse for om tiltaket har en slik ønsket effekt.

I denne rapporten er det gjort et tilfeldig utvalg på 10 % av alle personer som i løpet av tidsperioden fra 2002 til og med 2009 er oppgitt med minst et arbeidsforhold av minst 200 dagers varighet. Totalt var dette 283 296 personer. Av disse er alle med minst en sykmelding i samme tidsperiode fulgt i forløpet fra sin første sykmelding. Forløpet etter sykmelding er registrert som skiftene (overgangene) og vandringen mellom tilstandene *sykmelding*, *gradert sykmelding* og *jobb* for disse personene. Det var 219 438 personer (77.5 %) som hadde minst en sykmelding i den observerte tidsperioden. Totalt i forløpet var det for disse personene registrert 1 968 236 fulle sykmeldinger (83 %) og 398 843 graderte sykmeldinger (17 %).

Ved hjelp av utvidede statistiske modeller for forløpsanalyser (flertilstandsmodeller) følger vi personer mellom *full sykmelding*, *gradert sykmelding* og *jobb* og ved hvilke datoer personene skifter mellom disse tilstandene. Personene er fulgt i alle mulige overganger mellom disse tre tilstandene i løpet av tidsperioden fra 2002 og ut 2009. Basert på de observerte overgangene beregner vi; med hvilken hyppighet personer skifter mellom de tre tilstandene, med hvilken sannsynlighet de er i de ulike tilstandene, og antall dager de er i hver av de tre tilstandene. Analysene er gjort separat for ulike alderskategorier (under 40 år, 40-60 år og over 60 år) og separat for menn og kvinner.

Generelt fant vi at personer raskt kommer tilbake til *jobb*. De som har en *gradert sykmelding* har en lavere sannsynlighet for å komme tilbake til *jobb* enn de som er på *full sykmelding*. Dette gjaldt både for menn og kvinner separat og for de ulike kategorier av alder. Kvinner har høyere sykefravær enn menn. Vi viser at dette kan forklares ved at kvinner oftere skifter fra *jobb* til *sykmelding*, det vil si at de oftere blir sykmeldt men det ikke er en forskjell i lengden av fraværet mellom menn og kvinner når de først er sykmeldt. For alder så vi at eldre har høyere sykefravær. Vi forklarer dette med at eldre skifter senere fra *sykmelding* tilbake til *jobb*, det vil si at de ikke blir oftere sykmeldt, men når eldre først blir sykmeldt så er de sykmeldt lengre.

Ved å utnytte informasjonen som ligger i skifte mellom de tre tilstandene *full sykmelding*, *gradert sykmelding* og *jobb* og hvor lenge personer er i disse tilstandene i utvidede modeller for forløpsanalyser har vi synliggjort nye sider i prosessene for retur til jobb. Samtidig ivaretar vi egenskaper i dataene som er vanskelig å synliggjøre i andre analyser. For gradert sykmelding er det slik at de fleste ikke har dette som sin første sykmelding. Det vil si at et forløp fra gradert sykmelding ofte starter med en full sykmelding, før personen får en gradert sykmelding. Slike forløp ivaretas i de analysene vi har foretatt.

Vi fant at de som er innom gradert sykmelding har en lavere sannsynlighet for å returnere til jobb. Dette kan være fordi disse er en selektert gruppe individ med mer langvarige sykmeldinger og lidelser, og ikke en gruppe friskere individ med høyere arbeidsevne. Det kan også være slik at det å få en gradert sykmelding faktisk reduserer sannsynligheten for å returnere til jobb.

At kvinner har høyere sykefravær enn menn er kjent og flere årsaker til dette kan diskuteres. Vi har vist at forløpet som ligger bak forskjellen mellom menn og kvinner er at kvinner oftere blir sykmeldt enn menn, men når de først er sykmeldt er veien tilbake til jobb den samme.

Økende alder er også en kjent faktor for økende sykefravær. I denne rapporten viser vi videre at eldre ikke oftere blir sykmeldt, men når de først er sykmeldt er de det lengre enn yngre. En mulig forklaring på dette kan være at insitamentet for å holde seg i arbeid er konstant med alder, mens det med økende alder er en økende krav til tid for restituering.

Basert på våre enkle analyser kan det for kvinner kanskje være fornuftig å fokusere på tiltak som forhindrer at de *blir* sykmeldt, mens for eldre arbeidstagere kan det kanskje være fornuftig med tiltak for å få dem raskere tilbake til jobb når de først *er* sykmeldt.

## 1. Innledning

Bakgrunnen for denne rapporten er et undervisningsopplegg gjennomført for seksjonene Resultatoppfølging og analyse og Statistikk og utredning i Arbeids- og velferdsdirektoratet (NAV) i løpet av høsten 2009 og våren 2010. Undervisningsopplegget ble gjennomført i regi av Uni helse og hadde som formål å bidra til en kunnskapsheving av personell tilknyttet de to avdelingene. Først og fremst var dette knyttet til statistiske analyser av typen forløpsanalyser.

Forløpsanalyser er oftest knyttet til hendelser som skjer over tid. Tiden fra en person blir sykmeldt til personen returnerer til arbeid er en type hendelse som kan betraktes på denne måten. Hadde retur til arbeid vært en hendelse som skjedde kun en gang, altså at en person som er sykmeldt returnerer til arbeid og blir der for all fremtid (slik man kanskje kunne ønske), kunne man benyttet standard metoder for analyse av sykefravær. Men personer skifter ofte gjentatte ganger mellom arbeid og sykmeldt. I tillegg kan andre ytelser også være aktuelle. Denne relativt komplekse virkeligheten gjenspeiles i datastrukturene i NAV.

De mange ulike ytelser og gjentatte episoder for enkelte av ytelsene kan ivaretas i forløpsanalyser. Datastrukturen som ligger i databasene for de ulike ytelsene i NAV har flere likhetstrekk med det man kaller telleprosessdata. Det vil si at en person med flere ytelser har en linje for hver ytelse, slik at vi enkelt sagt kan telle datalinjer for å få vite antall ytelser enten for enkeltpersoner, grupper av personer eller totalt.

Denne rapporten beskriver resultater av forløpsanalyser av sykmeldingsdata, og om forløpsanalyser kan vise om gradert sykmelding påvirker sjansen for retur til arbeid. Vi har valgt en forholdsvis enkel flertilstandsmodell med de tre tilstandene *full sykmelding*, *gradert sykmelding* og *jobb* for å studere overgangene mellom disse og sannsynlighetene for å være i hver av disse tre tilstandene. I tillegg valgte vi å se på eventuelle forskjeller mellom menn og kvinner, ulike alderskategorier og ulike kategorier av kalenderår.

## 2. Datautvalg

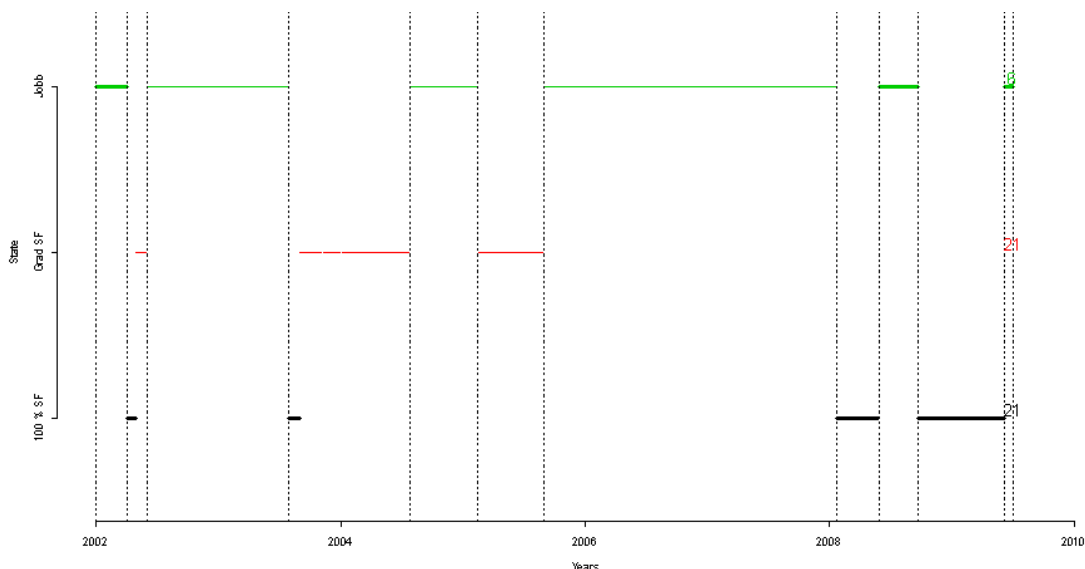
For analysene i denne rapporten er det gjort et tilfeldig datauttrekk av 10 % av alle som var registrert med minst et arbeidsforhold som varte i minst 200 dager sammenhengende i perioden 1. januar 2002 til utgangen av 2009. Totalt var det 283 296 personer i dette utvalget, 138 007 (48.7 %) kvinner. For disse personene var det registrert totalt 2 367 079 sykmeldinger. Kvinner hadde 1 394 964 (59 %) av antallet sykmeldinger. Fordeling av sykefravær etter kjønn og grad er vist i Tabell 1.

**Tabell 1: Fordeling av 100 % sykmelding og gradert sykmelding etter kjønn.**

	100 % Sykmeldt	Gradert (<100 %) sykmeldt	Totalt
<b>Kvinner</b>	1 117 920 (80 %)	277 044 (20 %)	<b>1 394 964</b>
<b>Menn</b>	850 316 (87 %)	121 799 (13 %)	<b>972 115</b>
<b>Totalt</b>	<b>1 968 236 (83 %)</b>	<b>398 843 (17 %)</b>	<b>2 367 079</b>

Hensikten med denne rapporten er å studere overgangen mellom full (100 %) sykmelding, gradert (<100 %) sykmelding og jobb. Vi ønsker derfor å følge enkeltindividene i forløpet og skiftene mellom disse 3 ulike tilstandene. Andre ytelser (rehabiliteringspenger, uførestønad osv.) er ikke med i betraktningen i analysene. Personer som er i slike ytelser vil da være utelatt (intervallsensurert) mens de eventuelt er på slike ytelser.

**Figur 2.1:** Eksempel på et individ og skift mellom jobb (grønn linje), gradert (<100 %) sykmelding (rød linje) og full (100 %) sykmelding. Denne personen er ikke fulgt helt til utgangen av 2009 siden det siste arbeidsforholdet (grønn linje avsluttes før dette). Denne personen er observert totalt med 11 graderte sykmeldinger og 21 fulle sykmeldinger, fra 2002 til 2009.



Siden vi ønsker å se på sannsynligheten og forløpet for at personer på gradert eller full sykmelding returnerer til arbeid er det ikke hensiktsmessig (eller riktig) å se på varigheten av enkelttilfeller. I mange tilfeller kan aggregerte størrelser, så som antall dager sykmeldt, antall sykmeldingsepisoder osv., være tilstrekkelig for å illustrere at det er eller kan være akkumulerte forskjeller mellom grupper (Hensing, Alexanderson et al. 1998). Når man ser på flere utfall samtidig blir dette mer problematisk, samtidig som man da mister den vesentlige

informasjonen om forløpet personene har. Vi har derfor valgt å definere og benytte sekvenser av sykmelding i våre analyser. Sykefraværsekvensene er definert ved start som en full (100 %) sykmelding eller gradert (<100 %) sykmelding inntil personen eventuelt er tilbake i jobb mer enn et år (365 dager) sammenhengende. Det er altså kun personer som har hatt minst en sykmelding i perioden 2002 til utgangen av 2009 som inngår i selve analysene. Personen i figur 1 vil i denne definisjonen ha 3 sekvenser av gradert sykmelding, full sykmelding og jobb. Den første starter sekvensen ved første fulle sykmelding i 2002, den neste starter med en full sykmelding i 2003 (etter at personen har vært mer enn 365 dager i jobb uten fravær, grønn linje), mens den siste sekvensen starter med en full sykmelding i 2008. Denne definisjonen og bruken av et fullt år i arbeid for å definere skille mellom sekvenser kan diskuteres, men vi mener at en slik struktur generelt evner å fange opp den vandring/skifting personer har mellom gradert sykmelding, full sykmelding og jobb på en grei måte.

Et annet alternativ kunne vært å benytte 6 måneder i jobb (som er kravet for å opparbeide full nye rettigheter til sykepenger), men da vil kanskje analysene påvirkes mer av denne regelen enn av den faktiske progresjonen personene har.

I tabell 2 er en oversikten for antall sekvenser for de 219 438 personene som inngår i analysene. For kalenderårene 2002 og ut 2009 er det altså slik at for de 10 % av befolkningen med minst et arbeidsforhold av minst 200 dagers varighet, totalt 283 296 personer, er det 219 438 personer (77.5 %) som har minst en sykmelding.

**Tabell 2:** Antall sekvenser av gjentatte graderte sykmeldinger, fulle sykmeldinger og jobb

<b>Antall sykefraværs</b>	
-	
<b>sekvenser</b>	<b>Personer</b>
1	103 035
2	71 642
3	34 524
4	9 202
5	1 004
6	31
<b>Totalt</b>	<b>219 438</b>

Som en illustrasjon på antall gjentakende ganger personer totalt skifter mellom de tre tilstandene og antall ganger for hver av de tre tilstandene for seg er det i tabell 3 gitt en oversikt over dette.

**Tabell 3:** Totalt antall tilstander personer har vært innom i løpet av tidsperioden, samt antall ganger de enkelte personer har vært i hver av de tre tilstandene.

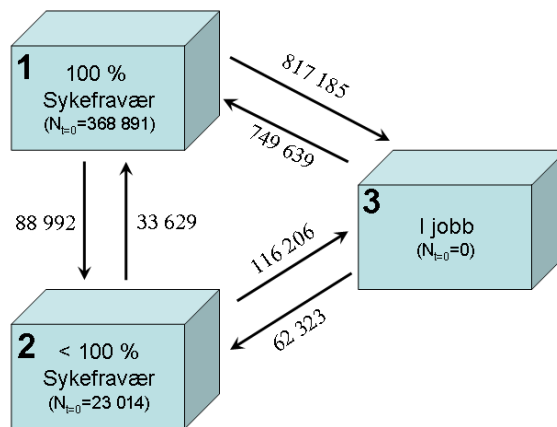
	<b>Totalt</b>	<b>100 % Sykmeldt</b>	<b>Gradert (&lt;100 %) sykmeldt</b>	<b>Jobb</b>
Antall personer	219438	217253	83569	214084
Antall ganger:				
- Gjennomsnitt	10.65	4.84	2.06	5.20
- Median	8	3	1	4
- Hyppigst forekommende	2	1	1	1
- Standard avvik	9.66	4.55	1.58	4.58
- Minimum	1	1	1	1
- Maksimum	187	93	51	92



## 2.1. Flertilstandsmodell

Vi definerte en 3-tilstandsmodell for å synliggjøre vandringen personer har mellom gradert sykmelding, full sykmelding og jobb. Denne modellen er visualisert i figur 2. Vi ser her at det er 368 891 sykefraværsekvenser (94 %) som starter med en full (100 %) sykmelding, mens det er 23 014 sekvenser (6 %) som starter med en gradert (<100 %) sykmelding. Videre ser vi at det åpenbart er flest skift (overganger) mellom jobb og full (100 %) sykmelding (749 639 og 817 185 observasjoner henholdsvis).

**Figur 2.2:** Oversikt over antall skift mellom de tre tilstandene [1] full (100 %) sykmelding, [2] gradert (<100 %) sykmelding og [3] jobb.



For å illustrere forløpet av sykmeldinger ytterligere har vi valgt ut et sett med sentrale og enkle forklaringsvariable. Variablene vi har sett på er alder (alder ved oppstart av sykefraværsekvensen), kjønn og år (kalenderår for oppstart av sykefraværsekvensen). I tabell 4 er det oppgitt frekvensfordeling for andel med gradert sykmelding for hver av kategoriene innenfor de tre variablene, når vi har definert flertilstandsmodellen som over. Tallene vil således også vise andel gradering generelt for de som har hatt sykmeldingsepisoder i perioden 2002 og ut 2009. Analysene er så gjort stratifisert for disse variablene (supplert med noen enkle regresjonsmodeller for overgangsintensitetene, delvis beskrevet i avsnitt 3. Statistisk metode).

**Tabell 4:** Andel med gradert sykmelding

		<b>Prosent gradert sykmelding totalt i forløpet</b>	<b>Prosent gradert sykmelding ved oppstart<sup>a</sup></b>
<b>Alder</b>	<b>Under 40 år</b>	11.9 % (663794)	5.2 % (193600)
	<b>40-60 år</b>	16.7 % (498248)	6.5 % (171075)
	<b>Over 60 år</b>	16.7 % (54713)	6.9 % (25308)
<b>År</b>	<b>2002-2004</b>	13.4 % (698767)	5.7 % (177520) <sup>b</sup>
	<b>2005-2007</b>	15.1 % (370492)	5.8 % (129501)
	<b>2008-2009</b>	14.5 % (154633)	6.3 % (84884)
<b>Kjønn</b>	<b>Kvinner</b>	16.6 % (719048)	7.7 % (208582)
	<b>Menn</b>	10.4 % (504844)	3.8 % (183323)
<b>Totalt</b>		<b>14.1% (1223892)</b>	<b>5.9% (391905)</b>

<sup>a</sup> Gradert sykmelding er først ved oppstart av en ny sykefraværsskvens.

<sup>b</sup> Tallet kan være upresist siden observasjonene starter 1.januar 2002, mens sykefraværperiodene kan ha startet tidligere.

### 3. Statistisk metode<sup>1</sup>

(Dette avsnittet er noe mer teknisk og man trenger ikke lese det for å få utbytte av rapporten.)

Statistiske analyser for utvidede forløpsdata og modeller for flertilstandsdata har hatt en vesentlig utvikling de senere årene (Commenges 1999; Andersen and Keiding 2002; Commenges 2002; Lie, Eriksen et al. 2008). Dette kapittelet oppsummerer noen av begrep og noen generelle formler innenfor statistiske analyser av flertilstandsmodeller.

#### 3.1. Overgangsintensiteter

Innen forløpsanalyser danner hasardfunksjonen (intensitetsfunksjonen, ofte skrevet som  $h(t)$  eller  $\lambda(t)$ ) svært ofte en basis. Hasardfunksjonen i seg selv er ikke en sannsynlighet, men man kan betrakte den som en sannsynlighet over et lite tidsintervall ( $dt$ ) ved formelen.

$$h(t) \cdot dt = P(t < T \leq t + dt \mid T > t) \quad (\text{formel 1})$$

Det vil si at hasarden uttrykker sannsynligheten for en hendelse i et lite tidsintervall ( $dt$ ) etter tidspunkt  $t$ , gitt at personen var observert under risiko ved tidspunkt  $t$ .

Hasardfunksjonen er blant annet sentral for en lang rekke regresjonsmodeller<sup>2</sup> innen forløpsanalyse.

I den enkle flertilstandsmodellen i figur 1, med 3 rekurrente (gjentakende) tilstander, beskrives overgangen mellom de tre tilstandene med hasardfunksjonene

$$h_{jk}(t) \cdot dt = P(X(t + dt) = k \mid X(t-) = j) \quad (\text{formel 2})$$

der  $j, k = 1, 2, 3$ . Definisjonen av hasardfunksjonene blir her intensiteten for at et individ skifter til tilstand  $k$  like etter tidspunkt  $t$  ( $t+dt$ ), gitt at personen var i tilstand  $j$  umiddelbart før (dvs  $t-$ ). Det vil si at observasjonen vår  $X$  er at personen er i tilstand  $j$  ved tid  $t-$ , og så observerer vi  $X$  at personen skifter til tilstand  $k$  ved tid  $t+dt$ .

Hovedpoenget med flertilstandsmodeller er at man ikke kan se på overgangene uavhengig av hverandre. Som figur 1 antyder kan for eksempel personer gå direkte fra 100 % sykefravær til jobb (og bli der), men de kan også gå fra 100 % sykefravær via gradert (<100 %) sykefravær til jobb – og de kan gjøre dette gjentatte ganger. For å ivareta at personer kan vandre flere veier og gjentatte ganger i modellen i figur 1 definerer vi en overgangsintensitetsmatrise  $\mathbf{h}(t)$ . I denne matrisen defineres rekkene hvilken tilstand individet kommer fra, mens kolonnene definerer hvilken tilstand individet går til. Diagonalelementene ( $h_{jk}(t)$ ,  $j=k$ ), som er negative, kan løselig tolkes som intensiteten for ikke å forlate tilstand  $j$ .

<sup>1</sup> Metoder og estimat for inferens (det vil si standardavvik, standardfeil, konfidensintervall og p-verdier) er utelatt. For informasjon om dette se for eksempel i bøkene av Andersen, P. K., B. Borgan, et al. (1993). *Statistical Models Based on Counting Processes*. New York: Springer-Verlag, Aalen, O. O., Ø. Borgan, et al. (2008). *Survival and Event History Analyses: A Process Point of View*. New York: Springer-Verlag.

<sup>2</sup> Hasardfunksjonen defineres blant annet i den mye brukte regresjonsmodellen for proporsjonale hasarder (ordinær Cox regresjon) Cox, D. R. (1972). Regression Models and Life-Tables. *Journal of the Royal Statistical Society Series B-Statistical Methodology*; 34(2): 187-&.. I modellen for proporsjonale hasarder antar man at hasardene for ulike kategorier av de inkluderte forklaringsvariablene ( $h(t|x)$ ) er proporsjonale (og multiplikative) over tid, mens den underliggende hasarden ( $h_0(t)$ ) er ikkeparametrisk. Cox modellen kalles derfor ofte semiparametrisk, siden den er en blanding av en ikkeparametrisk referansehasard og enkle parametre som beskriver forskjeller mellom kategorier av forklaringsvariable. Et alternativ til Cox modellen som kan nevnes er Aalens additive regresjonsmodell, som er en ren ikkeparametrisk modell der effektestimatene er funksjoner over tid (Aalen, O. O. (1989). A Linear-Regression Model for the Analysis of Life Times. *Statistics in Medicine*; 8(8): 907-925.)

$$\mathbf{h}(t) = \begin{pmatrix} - (h_{12}(t) + h_{13}(t)) & h_{12}(t) & h_{13}(t) \\ h_{21}(t) & - (h_{21}(t) + h_{23}(t)) & h_{23}(t) \\ h_{31}(t) & h_{32}(t) & - (h_{31}(t) + h_{32}(t)) \end{pmatrix} \quad (\text{formel 3})$$

I  $\mathbf{h}(t)$  vil hvert av elementene være funksjoner over tid, som ivaretar intensitetene for skifte mellom alle mulige tilstander. Estimeringen av de enkelte elementene i  $\mathbf{h}(t)$  kan gjøres på ulike måter (både parametrisk og ikke-parametrisk) og kan eventuelt også baseres på regresjonsmodeller, men vi går ikke videre inn på det her.

En standard metode for å beregne overgangsintensiteter, også i flertilstandsmodeller, er ved å benytte en ren ikkeparametrisk beregning av den integrerte (kumulerte) intensiteten,  $H_{jk}(t)$ , for hvert element, uavhengig av hverandre, ved Nelson-Aalen estimatoren. I flertilstandsmodeller kan Nelson-Aalen estimatoren defineres som

$$\hat{H}_{jk}(t) = \sum_{T_s \leq t} \frac{n_{jk}(T_s)}{Y_j(T_s)}. \quad (\text{formel 4})$$

Her er  $n_{jk}(t)$  antall skift (hendelser) fra tilstand  $j$  til tilstand  $k$  ved tidspunkt  $t$ , mens  $Y_j(t)$  er antall observasjoner i tilstand  $j$  ved tid  $(t)$ . Det vil si at  $Y_j(t)$  definerer risksettet for alle som kan skifte fra  $j$  til  $k$  ved tid  $t$ , men som ikke nødvendigvis gjør det.  $T_1 < T_2 < \dots$  er de sorterte observerte tidspunktene der det skjer en hendelse, det vil si skift fra  $j$  til  $k$ . Ved å beregne estimatene for  $H_{jk}(t)$  for alle  $j$  og  $k$  og arrangere dem i en kumulert (integrert) matrise for  $\mathbf{H}(t)$ , for overgangsintensitetene, ivaretar vi nå alle mulige overganger i modellen i figur 1.

### 3.2. Overgangssannsynligheter

Matrisen  $\mathbf{H}(t)$ , for overgangsintensitetene, kan i seg selv være av interesse, men sier ikke umiddelbart noe om sannsynlighetene for å være i de ulike tilstandene i figur 1. Vi ønsker nå et uttrykk for sannsynlighetene

$$P_{jk}(t_1, t_2) = P(X(t_2) = k | X(t_1) = j) \quad (\text{formel 5})$$

som er sannsynlighetene for at individet observeres ( $X$ ) i tilstand  $k$  ved tidspunkt  $t_2$ , gitt at individet var ( $X$ ) i tilstand  $j$  ved tidspunkt  $t_1$ . Denne sannsynligheten er da uavhengig av hvor (i hvilke tilstander) individet har vært mellom tidspunktene  $t_1$  og  $t_2$  og heller ikke hvor mange ganger individet har skiftet mellom de ulike tilstandene. Sannsynlighetene  $P_{jk}(t_1, t_2)$  kan nå settes opp i en overgangssannsynlighetsmatrise  $\mathbf{P}(t_1, t_2)$  definert som

$$\mathbf{P}(t_1, t_2) = \begin{pmatrix} P_{11}(t_1, t_2) & P_{12}(t_1, t_2) & P_{13}(t_1, t_2) \\ P_{21}(t_1, t_2) & P_{22}(t_1, t_2) & P_{23}(t_1, t_2) \\ P_{31}(t_1, t_2) & P_{32}(t_1, t_2) & P_{33}(t_1, t_2) \end{pmatrix}. \quad (\text{formel 6})$$

I overgangssannsynlighetsmatrisen vil (naturlig nok) summen av elementene i en rekke være 100 % ( $\sum_{k=1}^3 P_{jk}(t_1, t_2) = 1$ ). Det vil si at et individ i tilstand  $j$  ( $j=1, 2, 3$ ) ved tid  $t_1$  må være i en eller annen tilstand  $k$  ( $k=1, 2, 3$ ) ved tid  $t_2$ .

Matrisen for overgangssannsynligheter  $\mathbf{P}(t_1, t_2)$ , estimeres ofte ikkeparametrisk ved produktintegralet (Aalen-Johansen estimatoren) over inkrementene i de kumulerte intensitetsmatrisene der  $\Delta \hat{\mathbf{H}}(t)$  er det estimerte inkrementet fra den kumulerte intensitetsmatrisen (Aalen and Johansen 1978).

$$\hat{\mathbf{P}}(t_1, t_2) = \prod_{s=t_1}^{t_2} [\mathbf{I} + \Delta \hat{\mathbf{H}}(s)] \quad (\text{formel 7})$$

For enklere varianter av flertilstandsmodeller kan man analytisk gi uttrykk for de enkelte elementene i matrisen for sannsynligheter. For eksempel vil i den enklest mulige flertilstandsmodellen, med kun to tilstander (1 og 2) der tilstand 1 er transient og tilstand 2 er absorberende,  $P_{11}(0, t)$  være lik  $S(t)$  som er overlevelsessannsynligheten (sannsynligheten for å være i tilstand 1), som kan beregnes ved for eksempel Kaplan-Meier metoden. Videre vil det i den enkleste modellen være en en-entydig matematisk sammenheng mellom overlevelsessannsynligheten ( $S(t)$ ) og intensiteten ( $h(t)$ ) gitt som

$$S(t) = e^{-\int_0^t h(u) du} \quad (\text{formel 8})$$

Det vil si at hvis man vet  $h(t)$ , så vet man  $S(t)$  og vise versa. Fordelen med denne sammenhengen er åpenbar når man gjør regresjonsmodeller for intensiteten  $h(t)$  da dette også direkte vil si noe om sammenhengen for overlevelsessannsynligheten<sup>3</sup>. Den en-entydige sammenhengen man finner i den enkle modellen gjelder generelt ikke for generelle flertilstandsmodeller.

### 3.3. Tilstandssannsynligheter

Sannsynlighetsmatrisen  $\mathbf{P}(t_1, t_2)$  vil ofte være av spesiell interesse hvis man er interessert i de betingede sannsynlighetene, slik man er i denne rapporten. Det vil man for eksempel vite hva er sannsynligheten for å være i tilstanden "jobb" ved tid  $t_2$  gitt at personen var i tilstanden "100 % sykmeldt" ved tid  $t_1$ , samt sannsynligheten for å være i tilstanden "jobb" ved tid  $t_2$  gitt at personen var i tilstanden "gradert sykmeldt" ved tid  $t_1$ . Alternativt kan man være interessert i de ubetingede sannsynlighetene for å være i en gitt tilstand. Det vil si sannsynligheten  $Q$  for være i en gitt tilstand  $k$  ( $k=1, 2, 3$ ) ved tid  $t_2$  uavhengig av hvor individet var ved tidspunkt  $t_1$ . Noe forenklet definerer vi  $Q_k(t)$  som sannsynligheten for å være i tilstand  $k$  ved tidspunkt  $t$  uavhengig av hvor individet var ved tidspunkt 0.

$$Q_k(t) = P(X(t) = k) \quad (\text{formel 9})$$

$Q_j(0)$  er tilstandssannsynlighetene ved tidspunkt 0 ("nullfordelingen") oftest representert ved frekvensfordelingen ved tidspunkt 0. Tilstandssannsynlighetene  $Q_k(t)$  for  $k=1, 2, 3$  estimeres da ved

<sup>3</sup> For eksempel vil sammenhengen mellom overlevelsessannsynlighetene i en enkel proporsjonale hasarder modell der  $h_i(t|x) = h_0(t) \cdot e^{\beta \cdot x_i}$ , med én forklaringsvariabel  $X = \{0, 1\}$ , være  $S(t|X=1) = S(t|X=0)^{HRR}$ , der HRR er hasard rate ratioen gitt som  $e^\beta$ .

$$\hat{Q}_k(t) = \sum_{j=1}^3 \hat{Q}_j(0) \hat{P}_{jk}(0, t). \quad (\text{formel 10})$$

For enklere modeller, slik som den enkleste mulige nevnt ovenfor, men også modeller for konkurrerende årsaker (competing risks), og også andre modeller der alle individ starter i en og samme tilstand, vil  $Q(t)$  være lik  $P(0, t)$ . Dette er åpenbart siden  $Q_j(0)$  da er 0 for alle tilstander unntatt den ene tilstanden (J) der alle individ starter, der  $Q_j(0)=1$ . Ved komplett oppfølging for alle individ vil  $Q_k(t)$  være de observerte sannsynlighetene for de gitte tilstandene. Et eksempel på slike sannsynligheter er tidsseriene (der tid er kalendertid) for andel individ som er sykmeldt, for eksempel slik det publiseres av SSB (<http://www.ssb.no/sykefratot/>).

### 3.4. Tilstandsforventning

Videre kan man beregne forventet tid (gjennomsnittlig tid),  $E_k(t)$ , et individ oppholder seg i tilstand k fram til tid t, basert på tilstandssannsynlighetene  $Q_k(t)$ .

$$\hat{E}_k(t) = \int_0^t \hat{Q}_k(u) du \quad (\text{formel 11})$$

### 3.5. Regresjonsmodeller

For flertilstandsmodeller kan man selvsagt gjøre klassiske regresjonsmodeller på overgangsintensitetene  $h_{jk}(t)$ , for eksempel ved ordinær proporsjonale hasarder modeller (som i formel 12) stratifisert for hver av overgangene. Man kan også definere en kombinasjon av overganger eller felles parametre for hasarden for ulike overganger. Som en illustrasjon er det gjort enkle analyser for variabelen kjønn for de mulige overgangene i figur 1. Dette er ikke ment som en fullstendig analyse, men ment som en illustrasjon på hva regresjonsmodeller for overgangsintensitetene viser eller kan vise.

$$h_{i,jk}(t|x) = h_{0,jk}(t) \cdot e^{\beta_{jk}^T \cdot x_i} \quad (\text{formel 12})$$

For flertilstandsmodeller av typen konkurrerende årsaker kan man gjøre regresjonsmodeller direkte på tilstandssannsynlighetene (de kumulerte insidenser eller sub-hasardene) for å være i de ulike konkurrerende årsakene (Fine and Gray 1999; Scheike and Zhang 2007). For mer generelle flertilstandsmodeller genereres ofte pseudodata (ved jackknifing) for tilstandssannsynlighetene  $Q_{k(i)}(t)$  for hvert av individene, som så inngår i regresjonsmodeller (Andersen and Klein 2007) eller bruk av inverse sannsynlighetsvekter (Scheike and Zhang 2008). Dette diskuteres ikke videre i denne rapporten.

## 4. Resultater

Basert på modellen som er oppgitt i figur 2.2 beregner vi først de kumulerte (akkumulerte) intensiteten for skifte mellom de 3 ulike tilstandene (boksene); [1] full (100 %) sykmelding, [2] gradert (<100 %) sykmelding og [3] jobb i figur 4.1. Løselig sagt kan figurene for intensitetene forstås som den "hastighet" (den øyeblikkelige sjansen) personer skifter langs de markerte pilene i figur 2.2. Hvis for eksempel en person har vært sykmeldt i 30 dager, hva er da sjansen for at personen skifter tilbake til jobb dag 31? Stigningen på kurvene for intensitetene (siden kurvene er kumulerte) angir altså den øyeblikkelige sjansen for å forflytte seg mellom tilstandene i figur 2.2. En bratt kurve betyr høy øyeblikkelig sjanse (intensitet), mens en flat kurve betyr lav øyeblikkelig sjanse. Som en generell betraktning ser vi at kurvene for å gå fra full (100 %) sykmelding til jobb er svært bratt i begynnelsen. Kurvene for sannsynligheter viser med hvilken sannsynlighet personer *er* i de ulike tilstandene i figur 2.2. Det kan for eksempel være sannsynligheten for en person som starter som fullt sykmeldt å være i de tre ulike tilstandene; fortsatt fullt (100 %) sykmeldt, gradert sykmeldt eller i jobb. Sannsynlighetskurven for å være fullt sykmeldt vil da starte ved 100 % (siden personen startet som fullt sykmeldt) og vil generelt falle over tid, mens sannsynlighetene for å være gradert sykmeldt eller i jobb vil starte ved 0 % og generelt vil øke med tid.

### 4.1. Totale resultater

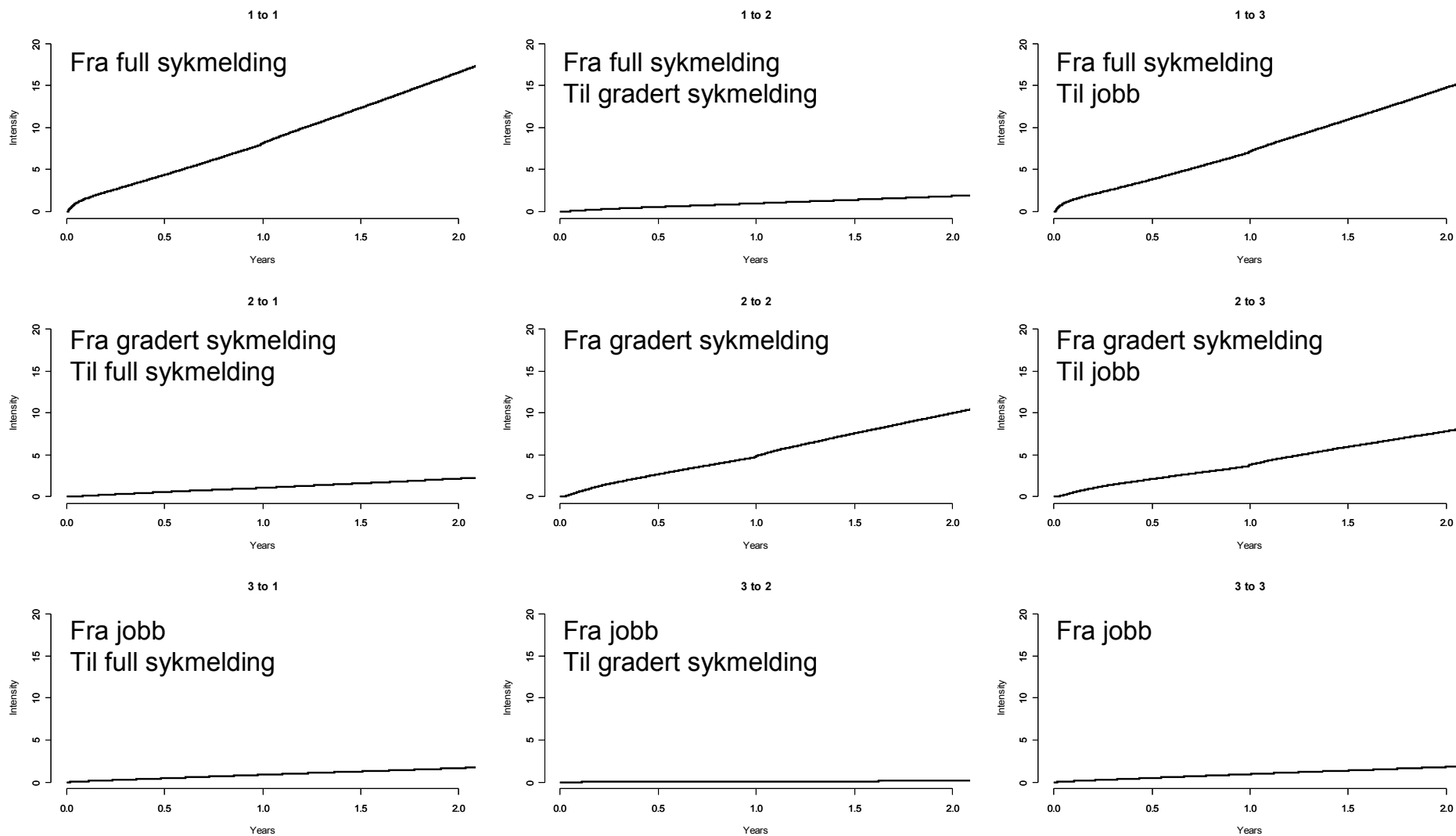
Diagonalelementene i figur 4.1 er sum av de øvrige elementene på samme rekke og angir den totale kumulerte intensiteten for å skifte fra den gitte tilstanden. Som vi ser i figuren er den høyeste intensiteten fra tilstand full (100 %) sykmelding til tilstand jobb og at den er høyest (brattest) helt i starten, mens den nest høyeste intensiteten er fra tilstand gradert (<100 %) sykmelding til tilstand jobb.

Selv om intensiteten for å skifte fra full sykmelding til jobb er dobbelt så høy som intensiteten for å skifte fra gradert sykmelding til jobb (hasard rate ratio HRR=2.0,  $p<0.001$ ) vil det ikke nødvendigvis si det samme som at sannsynligheten for skifte til jobb fra full (100 %) sykmelding er dobbelt så stor som å skifte til jobb fra gradert (<100 %) sykmelding.

Sannsynligheten for å være i jobb hvis du er fullt (100 %) sykmeldt er avhengig av flere faktorer. Blant annet vil hvor lenge personene er på full (100 %) sykmelding før de skifter til jobb (intensiteten) bety noe, men også den motsatte retningen, det vil si hvor lenge personer er i jobb før de igjen blir fullt (100 %) sykmeldt er avgjørende. Sannsynligheten for å være i jobb er da intuitivt avhengig av "summen" av de to retningene. Videre kan personer gå "omveien" fra full (100 %) sykmelding via gradert (<100 %) sykmelding tilbake til jobb, og motsatt retning tilbake til full (100 %) sykmelding. Personer kan altså gå gjentatte ganger langs alle piler i figur 2.2. "Størrelsen" på pilene uttrykkes ved intensitetene, mens sannsynlighetene for å være i de ulike tilstandene er "summen" av alle mulige kombinasjoner av retninger i figur 2.2 foran.

Vi observerer at intensitetene for skifte fra tilstand jobb til full (100 %) sykmelding og fra jobb til gradert (<100 %) sykmelding er lavere enn de andre intensitetene, der den sistnevnte er lavest.

**Figur 4.1:** Kumulerte intensiteter ( $H(t)$ ) for å skifte mellom de tre ulike tilstandene [1] full (100 %) sykmelding, [2] gradert (<100 %) sykmelding og [3] jobb. (95 % konfidensintervall faller sammen med de beregnede kurvene). For alle kurvene viser x-aksene oppfølgingstid fra 0 til 2 år, mens skalaen på y-aksene går fra 0 til 20 for å visualisere at det er ulikheter i størrelsene på intensitetene i figur 2.2.





Beregning av sannsynlighetene for å være i de tre ulike tilstandene i figur 2.2 baserer seg på de beregnede kumulerte intensitetene i figur 4.1.

Tilbake til jobb fra full sykmelding har, som illustrert i figur 2.2, flere muligheter. Personen kan for eksempel returnere direkte til jobb – og bli der, eller gå fra full (100 %) sykmelding til gradert (<100 %) sykmelding til jobb og så videre – gjentatte ganger. Muligheten for gjentatt skifting mellom de ulike tilstandene er ivaretatt i den kumulerte intensitetsmatrisen i figur 4.1 og alle de kumulerte intensitetene danner derfor basis for beregning av sannsynlighetene for å være i de 3 gitte tilstandene. Hvis utgangspunktet er den tilstanden personene er i ved start av de definerte sykefraværsskvensene kan vi beregne sannsynlighetene for at personene er i hver av de tre tilstandene i figur 2.2 ved alle senere tidspunkt. Dette gjøres altså basert på alle mulige skift (overganger) i flertilstandsmodellen. Siden det kun er personer i tilstand full (100 %) sykmelding og tilstand gradert (<100 %) sykmelding ved oppstart av sykefraværsskvensene (siden dette var definisjonen på start av skvensene) beregnes det først kun sannsynligheter for å skifte *fra* disse to tilstandene.

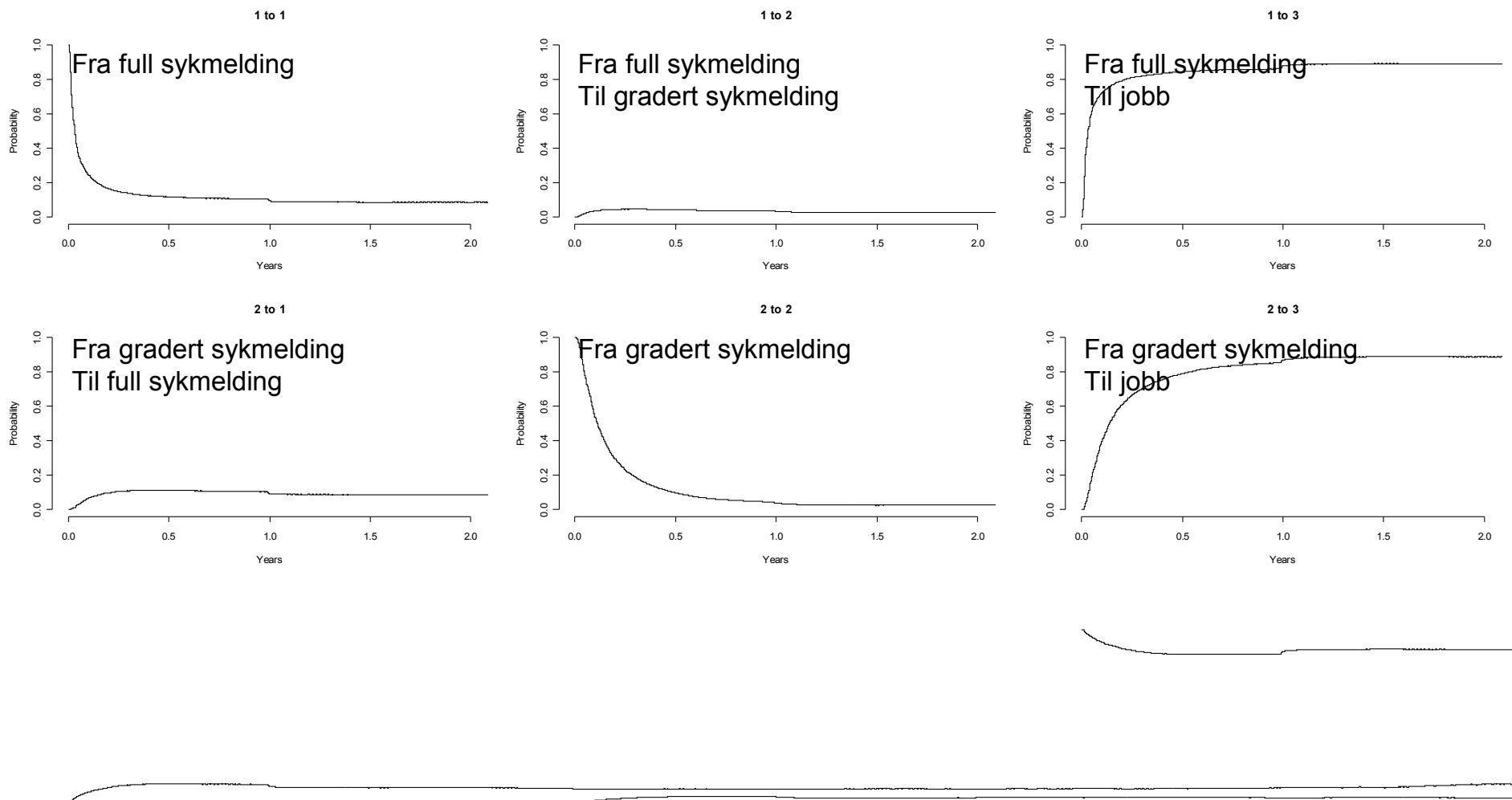
Ser vi på sannsynlighetene i figur 4.2 er det antagelig sannsynlighetene for å skifte fra tilstand full (100 %) sykmelding og fra tilstand gradert (<100 %) sykmelding til tilstand jobb av særlig interesse (det vil si ”Fra 1 til 3” og ”Fra 2 til 3”). Vi ser at sannsynligheten for å være i jobb ved et hvilket som helst tidspunkt er lavere hvis man starter sykefraværsskvensen med å være på gradert sykmelding. Forskjellen mellom de kurvene (til høyre i figur 4.2) kan også uttrykkes ved at personer som er gradert sykmeldt fra start av sykefraværsskvensen er 37 dager mindre i jobb i løpet av de 2 neste årene (og da først og fremst i løpet av det første året) enn personer som starter på full (100 %) sykmelding.

Siden andel på gradert (<100 %) sykmelding er lavere ved oppstart av sykefraværsskvensene enn totalt (tabell 2) har vi valgt å beregne sannsynligheter for å være i de ulike tilstandene gitt (betinget) på hvilken tilstand man var i ved 30 dager etter start av sykefraværsskvensen. Vi valgte 30 dager (1 måned) for å tydeligere effekten av det å være i en av de gitte tilstander når personene er over den første (akutte) perioden. Analysene er også gjort for 90 dager (3 måneder) og 182 dager (6 måneder), men de presenteres ikke siden de viser det samme som analysene for 30 dager. Ved 30 dager er det også personer i tilstand jobb, derfor beregnes også sannsynlighetene for å være *fra* denne tilstanden. Igjen er sannsynlighetene for å skifte fra tilstand full (100 %) sykmelding og fra tilstand gradert (<100 %) sykmelding til tilstand jobb av særlig interesse (det vil si ”Fra 1 til 3” og ”Fra 2 til 3”) av særlig interesse.

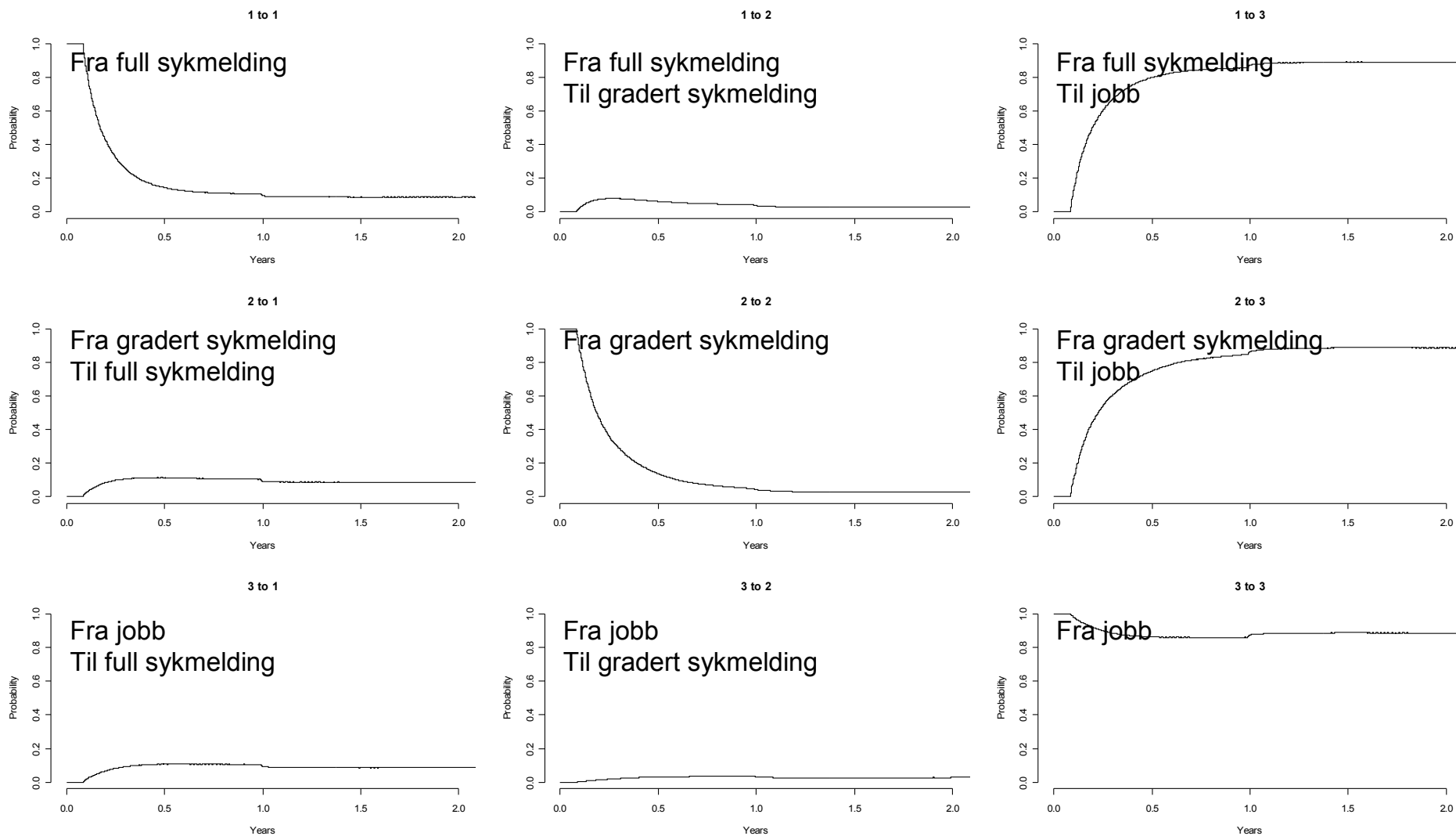
Også når vi utelukker de første 30 dagene i sykefraværsskvensene av analysene finner vi at sannsynligheten for å være i jobb er lavere for de som er på en gradert (<100 %) sykmelding. Totalt representerer dette her 13 dager mindre i jobb i løpet av de 2 neste årene for de som er gradert sykmeldt (og da først og fremst i løpet av det første året) enn personer som starter på full (100 %) sykmelding.

Til sist har vi også sett på sannsynlighetene for de ulike overgangene betinget på tilstandene ved 6 måneder (182 dager) og finner da også at sannsynligheten for å være i jobb er lavere for de som er på en gradert (<100 %) sykmelding enn de som er fullt (100 %) sykmeldt.

**Figur 4.2:** Sannsynligheter for å være i hver av de tre ulike tilstandene [1] full (100 %) sykmelding, [2] gradert (<100 %) sykmelding og [3] jobb, gitt (betinget) hvilken tilstand personen var i ved start av sykefraværsekvensen. For alle kurvene viser x-aksene oppfølgingstid fra 0 til 2 år, mens skalaen på y-aksene viser sannsynligheten (fra 0 til 1; 0 % til 100 %) for å gå *til* en gitt tilstand gitt hvilken tilstand man kom *fra*.



**Figur 4.3:** Sannsynligheter for å være i hver av de tre ulike tilstandene [1] full (100 %) sykmelding, [2] gradert (<100 %) sykmelding og [3] jobb, gitt (betinget) hvilken tilstand personen var i 30 dager etter start av sykefraværsekvensen. For alle kurvene viser x-aksene oppfølgingstid fra 0 til 2 år, mens skalaen på y-aksene viser sannsynligheten (fra 0 til 1; 0 % til 100 %) for å gå *til* en gitt tilstand gitt hvilken tilstand man kom *fra*.



## 4.2. Resultat for kjønn

Det er kjent at kvinner har høyere sykefravær enn menn (se for eksempel <http://www.nav.no/234017.cms>). Årsakene til at kvinner har høyere fravær kan være flere både private forhold, sykdom og arbeidsliv uten at vi skal gå videre inn på det her. I flertilstandsmodellene for kjønn er analysene gjort stratifisert, det vil si for menn og kvinner uavhengig av hverandre (med unntak av de enkle regresjonsmodellene). Hvis vi først ser på de beregnede kumulerte intensitetene i figur 4.3 ser vi at kurvene for menn og kvinner er identiske (sammenfallende) for overgangen fra full (100 %) sykmelding til jobb, også synliggjort ved en hasard rate ratio (HRR) lik 1.00. Også fra gradert (<100 %) sykmelding til jobb er det en ubetydelig forskjell. Derimot er det en høyere intensitet for kvinner enn for menn fra jobb (det vil si "Fra 3 til 1" og "Fra 3 til 2"). Det er også en høyere intensitet for kvinner enn for menn å skifte (begge veier) mellom gradert (<100 %) sykmeldt og full (100 %) sykmelding.

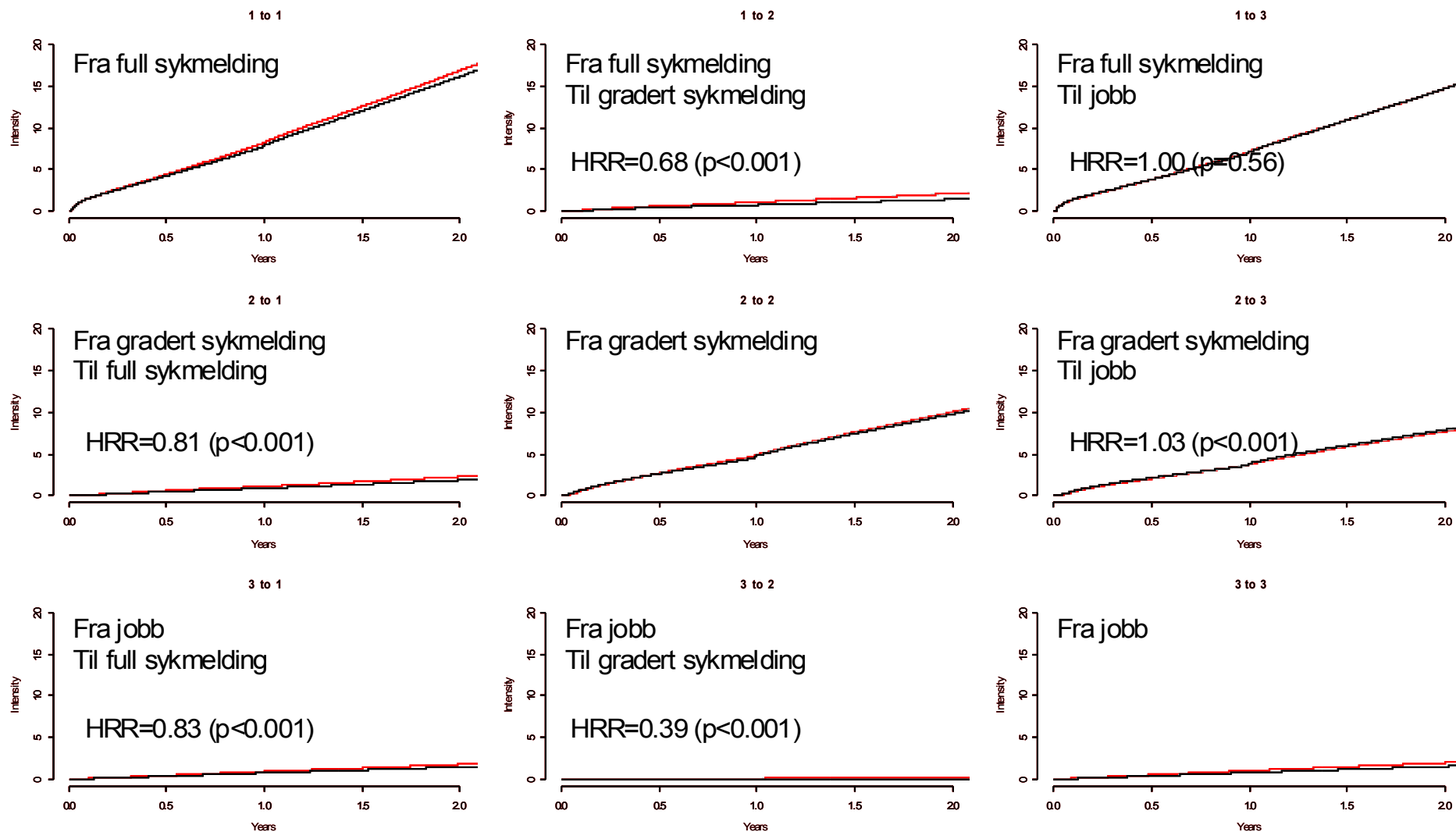
Basert på de ulike beregnede intensitetene for menn og kvinner beregner vi så de betingede sannsynlighetene for å være i de tre tilstandene i figur 2.2 for menn og kvinner uavhengig. I figur 4.5 ser vi for eksempel at sannsynligheten for at kvinner er i jobb gitt at hun var full (100 %) sykmeldt er lavere enn for menn ("Fra 1 til 3") og inverst at sannsynligheten for å være sykmeldt er høyere for kvinner enn for menn ("Fra 1 til 1") (NB! Selv om kurvene visuelt er nærme hverandre er dette grunnet skalaen på y-aksen. Nivået på tallene (prosentene) korresponderer med de offisielle oversiktene for ulikheter mellom kjønn fra SSB).

Som et supplerende mål for å synliggjøre forskjellen i sykefravær mellom menn og kvinner har vi beregnet forventet tid i de ulike tilstandene (formel 11). Tallene gjenspeiler gjennomsnittlig tid en mann respektivt kvinne er i de ulike tilstandene full (100 %) sykmelding, gradert (<100 %) sykmelding og jobb i 730 dager (2 år) etter første sykmelding (gradert eller full) i en ny sykefraværsekvens. Totalt estimerer vi da at menn vil ha vært 626 dager i jobb, mens kvinner har vært 604 dager i jobb. Vi kan forvente at menn i gjennomsnitt er 22 dager mer i jobb enn kvinner 2 år etter første sykmelding (tabell 5).

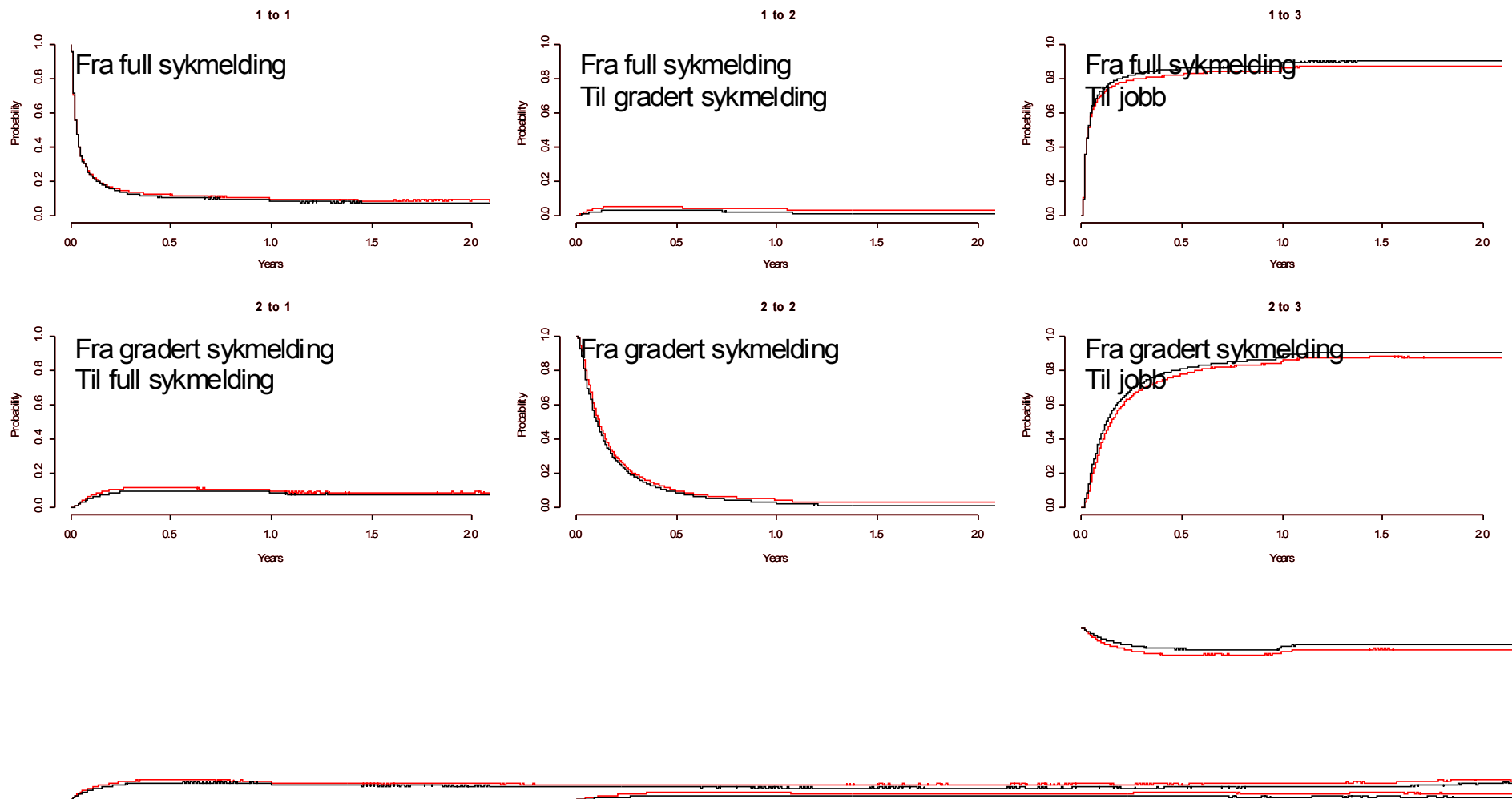
**Tabell 5:** Antall dager i de tre ulike tilstandene full (100 %) sykmelding, gradert (<100 %) sykmelding og jobb i løpet av 2 år (730 dager), for menn og kvinner, etter personen starter på en sykefraværsekvens, samt differansen i antall dager mellom menn og kvinner.

	100 % Sykmeldt	Gradert (<100 %) sykmeldt	Jobb
<b>Menn</b>	85 dager	19 dager	626 dager
<b>Kvinner</b>	92 dager	34 dager	604 dager
<b>Differanse</b>	-7 dager	-15 dager	22 dager

**Figur 4.4:** Kumulerte intensiteter (H(t)) for å skifte mellom de tre ulike tilstandene [1] full (100 %) sykmelding, [2] gradert (<100 %) sykmelding og [3] jobb. (95 % konfidensintervall faller sammen med de beregnede kurvene) for menn (sorte linjer) og kvinner (røde linjer). For alle kurvene viser x-aksene oppfølgingstid fra 0 til 2 år, mens skalaen på y-aksene går fra 0 til 20.



**Figur 4.5:** Sannsynligheter for å være i hver av de tre ulike tilstandene [1] full (100 %) sykmelding, [2] gradert (<100 %) sykmelding og [3] jobb, gitt (betinget) hvilken tilstand personen var i ved start av sykefraværsekvensen, for menn (sorte linjer) og kvinner (røde linjer). For alle kurvene viser x-aksene oppfølgingstid fra 0 til 2 år, mens skalaen på y-aksene viser sannsynligheten (fra 0 til 1; 0 % til 100 %) for å gå *til* en gitt tilstand gitt hvilken tilstand man kom *fra*.

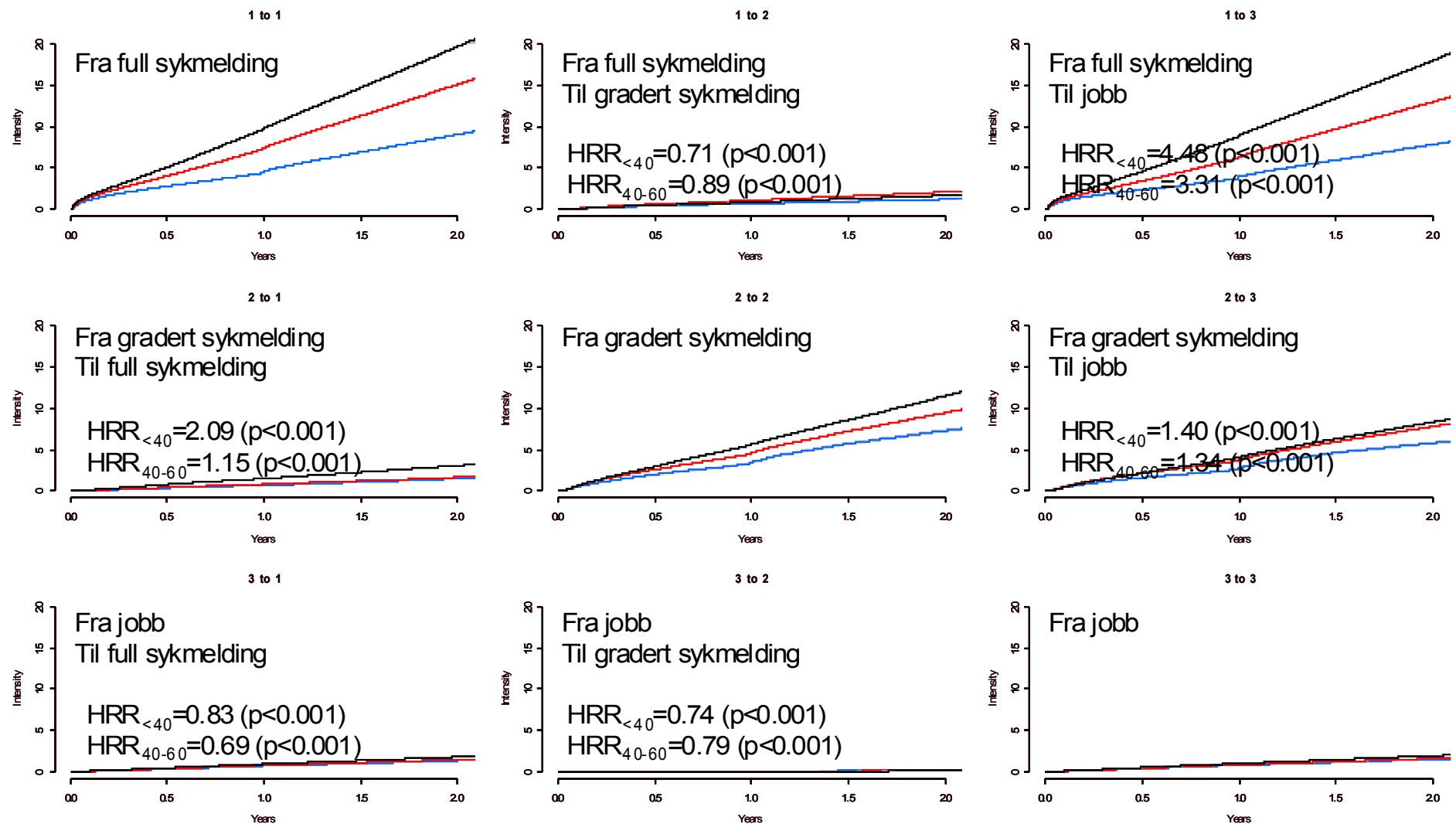


### **4.3. Resultat for alder**

Sykefravær øker med økende alder (se <http://www.nav.no/234017.cms>). Som for ulikhetene i kjønn kan man tenke seg flere forklaringsmodeller for hvorfor det er slik uten at vi skal diskutere det her. Vi for å diskutere og visualisere ulikheter i sykefravær for alder har vi valgt å dele alder inn i de tre kategoriene personer under 40 år (<40år), personer fra 40 til 60 år (40-60år) og personer over 60 år. Selv om dette er en grov inndeling mener vi det godt illustrerer hvorledes prosessen for retur til arbeid endres med økende alder. I figur 4.6 er de beregnede kumulerte intensitetene for de tre alderskategoriene. Vi legger spesielt merke til at det er stor forskjell mellom intensitetene for retur fra full (100 %) sykmelding til jobb for de tre alderskategoriene. Den yngste alderskategorien har en vesentlig høyere intensitet for å returnere til arbeid enn den eldste kategorien, mens den midtre kategorien ligger mellom disse. Dette kommer også tydelig fram i de enkle modellene for proporsjonale hasarder, der de yngste har en HRR=4.48 for retur til arbeid sammenlignet med de eldste. Hvis vi ser på intensitetene for å gå i motsatt retning, det vil si fra jobb til full sykmelding er ikke forskjellene så uthalt. Hasard rate ratioer (HRR), der personer over 60 år er referansegruppe er oppgitt i figur 4.6. Vi observerer også at yngre har en høyere intensitet for å skifte fra gradert (<100 %) sykmelding til full sykmelding, mens de har en lavere intensitet for å skifte i motsatt retning. Det vil si at forskjellene i skifte mellom de ulike tilstandene i figur 2.2 for alder er relativt komplekse.

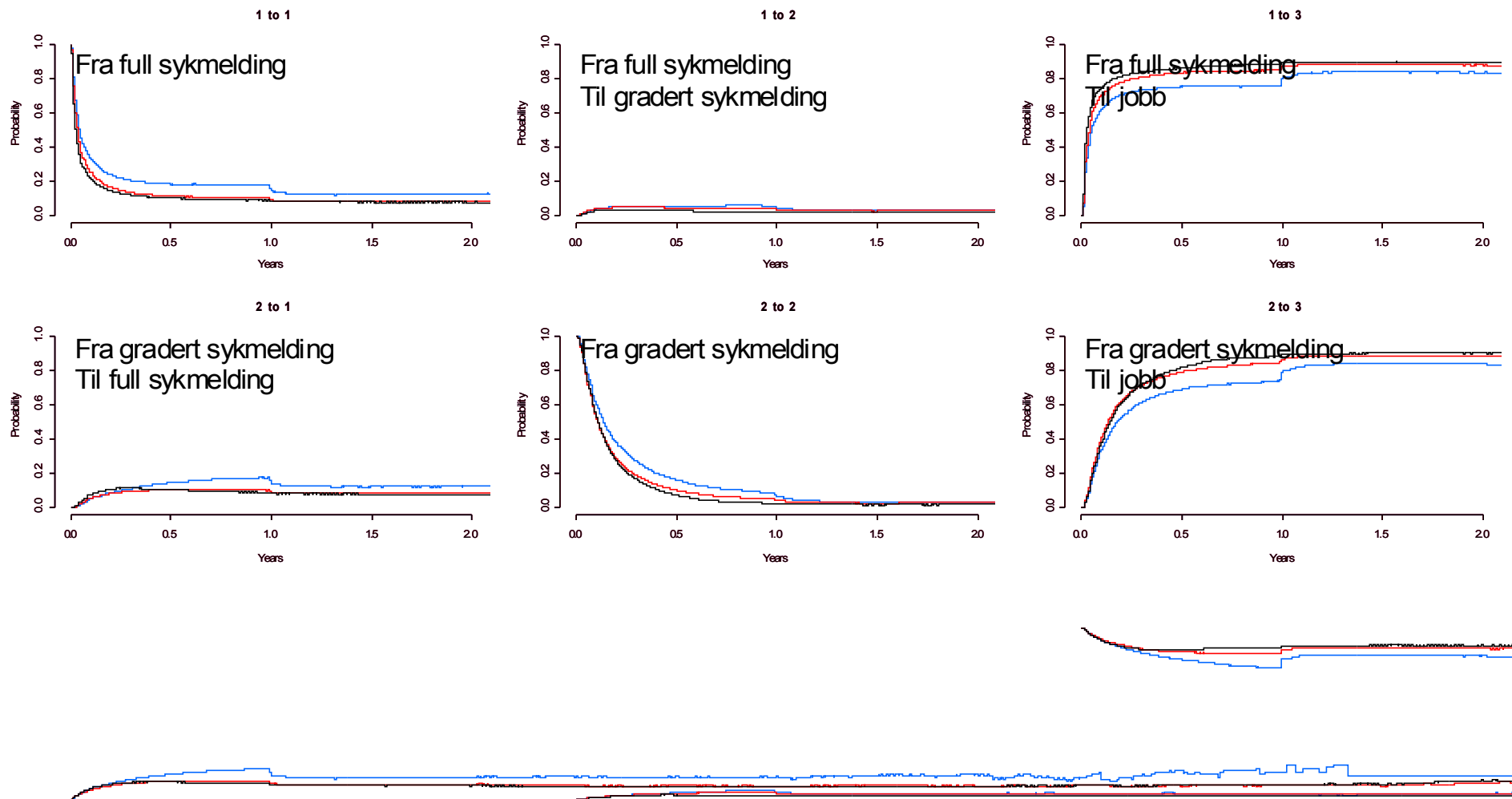
Akkumulert gir ulikhetene i intensiteter for alder i figur 4.6 også ulikheter i sannsynlighetene for å være i de ulike tilstandene i figur 2.2, særlig siden det var svært store ulikheter i intensitetene for å gå fra full (100 %) sykmelding til jobb. Som vi ser i figur 4.7 er sannsynlighetene for at personer med økende alder er tilbake i jobb, hvis de var fullt (100 %) sykmeldt eller gradert (<100 %) sykmeldt avtagende. Det vil si at med økende alder er det en økende sannsynlighet for å være sykmeldt (fullt eller gradert), det er kjent, og årsaken er at det først og fremst er ulikheter for de ulike aldersgruppene i intensiteten for å gå tilbake til jobb men ikke i motsatt retning.

**Figur 4.6:** Kumulerte intensiteter (H(t)) for å skifte mellom de tre ulike tilstandene [1] full (100 %) sykmelding, [2] gradert (<100 %) sykmelding og [3] jobb. (95 % konfidensintervall faller sammen med de beregnede kurvene) for personer under 40 år (sorte linjer), personer fra 40 til 60 år (røde linjer) og personer fra 60 år (blå linjer). For alle kurvene viser x-aksene oppfølgingstid fra 0 til 2 år, mens skalaen på y-aksene går fra 0 til 20.





**Figur 4.7:** Sannsynligheter for å være i hver av de tre ulike tilstandene [1] full (100 %) sykmelding, [2] gradert (<100 %) sykmelding og [3] jobb, gitt (betinget) hvilken tilstand personen var i ved start av sykefraværsekvensen, for personer under 40 år (sorte linjer), personer fra 40 til 60 år (røde linjer) og personer fra 60 år (blå linjer). For alle kurvene viser x-aksene oppfølgingstid fra 0 til 2 år, mens skalaen på y-aksene viser sannsynligheten (fra 0 til 1; 0 % til 100 %) for å gå *til* en gitt tilstand gitt hvilken tilstand man kom *fra*.

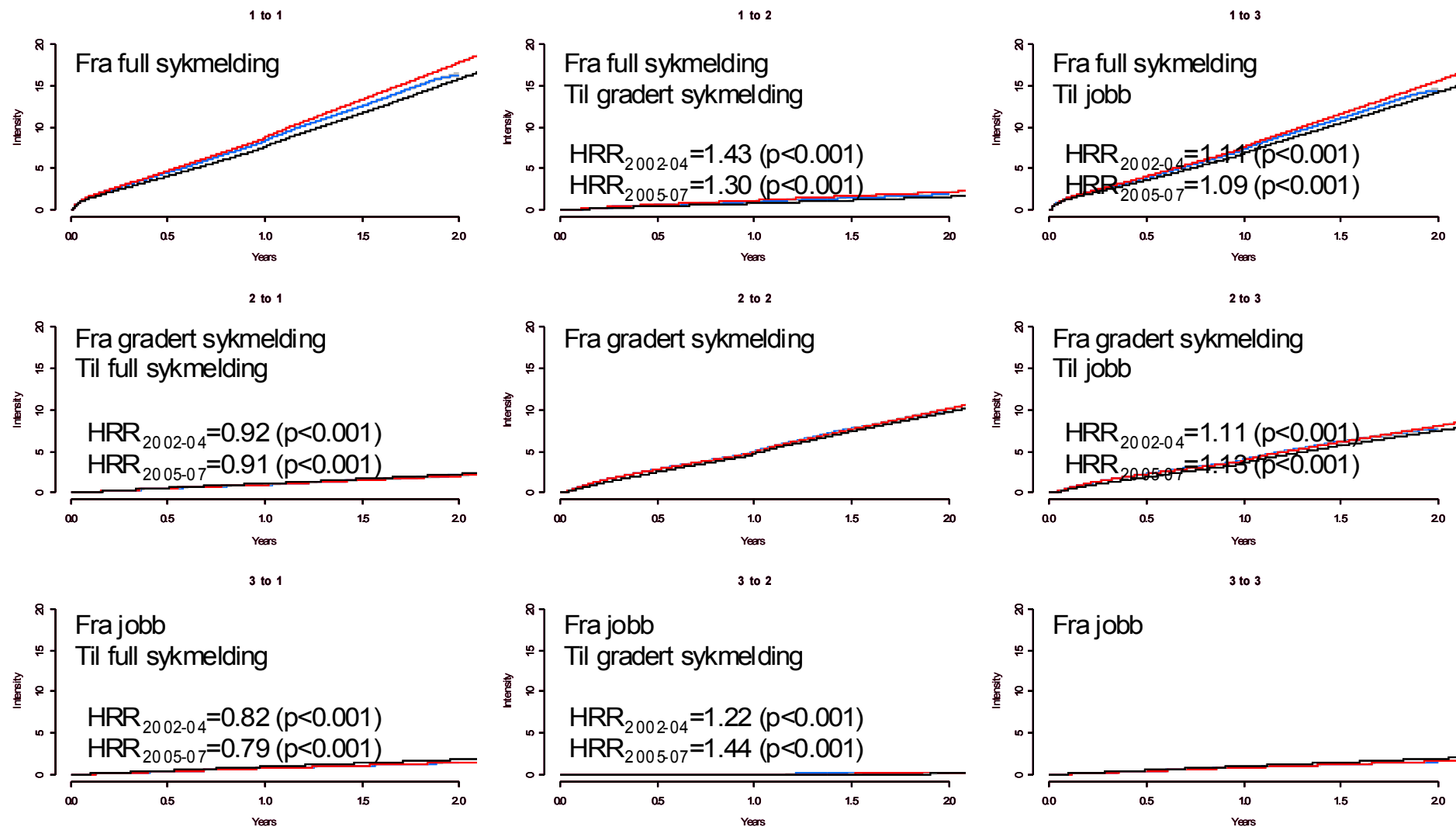


#### **4.4. Resultat for kalenderår**

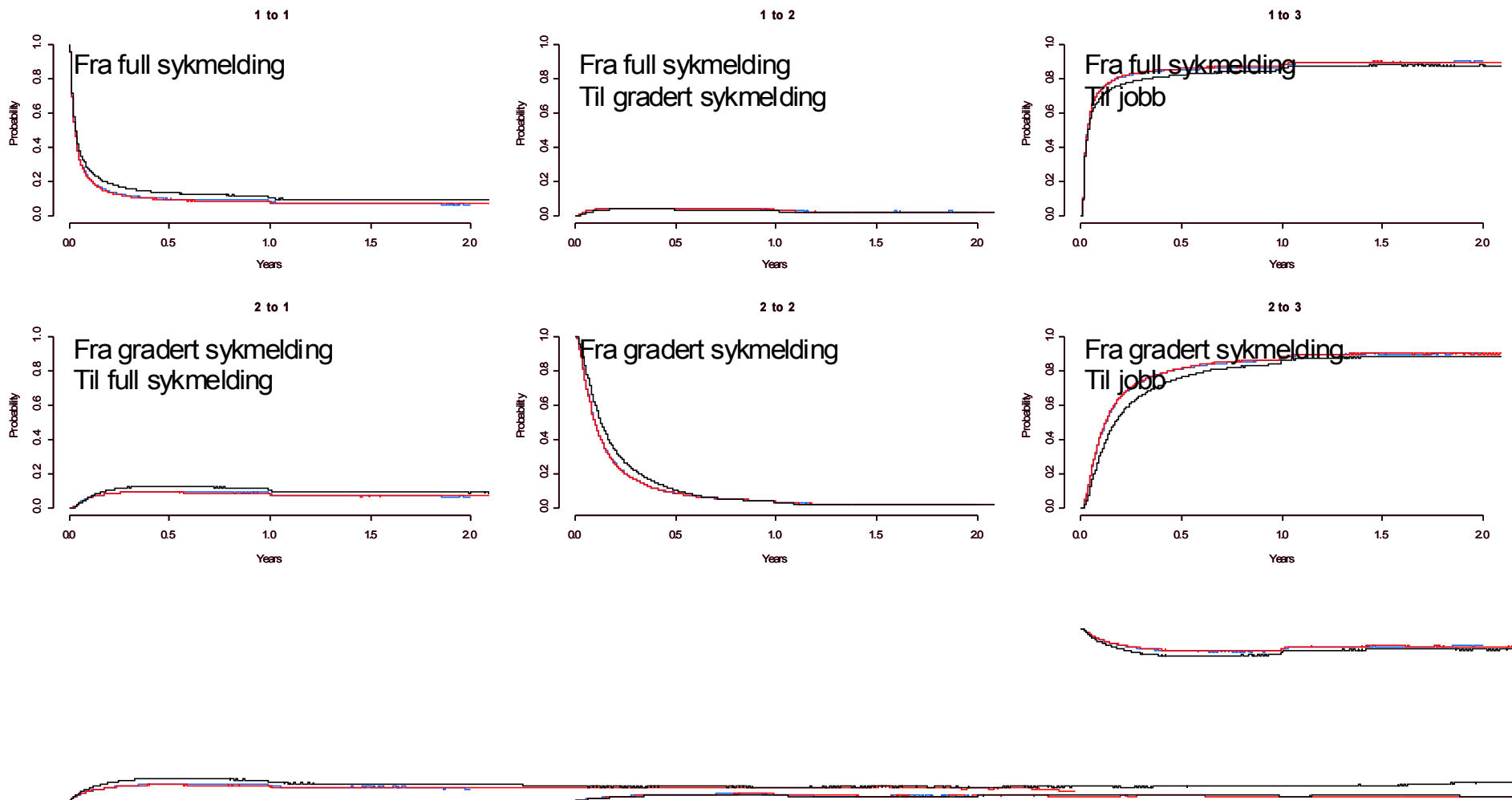
Endringer i sykmelding over kalenderår rapporteres vanligvis som tidsserier over andel sykmeldte, gjerne beregnet som andel tapte dagsverk (<http://www.ssb.no/sykefratot/>). En slik oversikt er hensiktsmessig for å se på generelle endringer over tid, samt at den er relativt enkel å beregne. Hasardfunksjoner kan beregnes der kalendertid er tidsskalaen. Det har vi ikke gjort her, men fortsatt holdt oss til de definerte sykefraværsekvensene der vi følger personer i 2 år etter start av en slik sekvens. For å se på eventuelle forskjeller for kalenderår er det laget tre kategorier; årene 2002-2004, årene 2005-2007 og årene 2008-2009. Dette er igjen kanskje en grov inndeling, men er mest ment for å illustrere eventuelle forskjeller mellom de tre tidsperiodene. I figur 4.8 er de beregnede overgangensintensitetene for de tre tidsperiodene. Det er en viss forskjell i overgangensintensitetene mellom de tre tilstandene i figur 2.2 for de tre tidsperiodene. Vi ser blant annet at det totalt er en lavere intensitet fra full sykmelding til jobb for den siste perioden sammenlignet med de to foregående tidsperiodene, mens det i motsatt retning ser ut til at det er en høyere intensitet for å i motsatt retning. Det er også en lavere intensitet i den siste tidsperioden å gå til gradert (100 %) sykmelding både fra full (100 %) sykmelding og fra jobb. De oppgitte hasard rate ratioene (HRR) i figur 4.8 er beregnet med den siste tidsperioden (2008-2009) som referansekategori.

Basert på de beregnede intensitetene i figur 4.8 finner vi at sannsynlighetene for å være i de tre ulike tilstandene (gitt om man startet sykefraværsekvensen med full (100 %) sykmelding eller gradert (<100 %) sykmelding) er lik for de to første tidsperiodene. Det vil si at for personer som blir sykmeldt i de to første tidsperiodene (2002-2004 og 2005-2007) er sannsynligheten for å være i jobb like. For den siste tidsperioden er sannsynlighetene for å være i jobb lavere enn for de to første periodene, mens sannsynlighetene for å være på full (100 %) sykmelding eller gradert (<100 %) sykmelding er høyere.

**Figur 4.8:** Kumulerte intensiteter ( $H(t)$ ) for å skifte mellom de tre ulike tilstandene [1] full (100 %) sykmelding, [2] gradert (<100 %) sykmelding og [3] jobb. (95 % konfidensintervall faller sammen med de beregnede kurvene) for årene 2002-2004 (sorte linjer), 2005-2007 (røde linjer) og 2008-2009 (blå linjer). For alle kurvene viser x-aksene oppfølgingstid fra 0 til 2 år, mens skalaen på y-aksene går fra 0 til 20.



**Figur 4.9:** Sannsynligheter for å være i hver av de tre ulike tilstandene [1] full (100 %) sykmelding, [2] gradert (<100 %) sykmelding og [3] jobb, gitt (betinget) hvilken tilstand personen var i ved start av sykefraværsekvensen, for årene 2002-2004 (sorte linjer), 2005-2007 (røde linjer) og 2008-2009 (blå linjer). For alle kurvene viser x-aksene oppfølgingstid fra 0 til 2 år, mens skalaen på y-aksene viser sannsynligheten (fra 0 til 1; 0 % til 100 %) for å gå *til* en gitt tilstand gitt hvilken tilstand man kom *fra*.



#### 4.5. Oppsummering

Siden det kan være vanskelig å vurdere de ulike figurene for sannsynligheter for de ulike overgangene i figur 2.2 (det vil si figurene 4.2, 4.3, 4.5, 4.7 og 4.9) er sannsynlighetene for å være i jobb ved et gitt tidspunkt gitt om man var *gradert sykmeldt* eller *fullt sykmeldt* ved 1 måned og ½ år beregnet og satt opp i tabell 6. Legg merke til at det her ikke er snakk om å være ”tilbake i jobb”, siden det kan oppfattes som om man da blir værende i jobb uten å bli sykmeldt igjen men at det i praksis er fullt mulig for personer å være sykmeldt både dagen før og dagen den oppgitte dagen. Tabellen gir en oversikt for å være jobb avhengig av om man er på gradert sykmelding eller full sykmelding for de ulike kategoriene av dataene og totalt.

**Tabell 6:** Sannsynligheter for å være i *jobb* ved 1 måned og ½ år gitt om man var på *full sykmelding* eller *gradert sykmelding* ved starten av en sykefraværsekvens.

<b>Til:</b>		<b>1 måned</b>		<b>½ år</b>	
<b>Jobb</b>		<b>Full sykmeldin g</b>	<b>Gradert sykmeldin g</b>	<b>Full sykmeldin g</b>	<b>Gradert sykmeldin g</b>
<b>Fra:</b>					
<b>Alder</b>	<b>Under 40 år</b>	72.7 %	31.7 %	86.3 %	81.5 %
	<b>40-60 år</b>	67.3 %	34.7 %	83.6 %	79.0 %
	<b>Over 60 år</b>	59.6 %	29.0 %	75.4 %	68.8 %
<b>År</b>	<b>2002-2004</b>	66.9 %	27.0 %	82.4 %	76.1 %
	<b>2005-2007</b>	71.4 %	37.4 %	85.9 %	81.3 %
	<b>2008-2009</b>	71.2 %	37.0 %	85.6 %	81.2 %
<b>Kjønn</b>	<b>Kvinner</b>	68.3 %	31.0 %	82.7 %	77.6 %
	<b>Menn</b>	70.5 %	36.9 %	86.0 %	80.7 %
	<b>Totalt</b>	69.3 %	32.9 %	84.2 %	79.0 %
	<b>Totalt<sub>1 måned</sub><sup>a</sup></b>	41.4 %	36.5 %	82.0 %	78.1 %
	<b>Totalt<sub>1/2 år</sub><sup>b</sup></b>	38.0 %	23.0 %	85.1 %	77.4 %

a Gitt at de var på full sykmelding eller gradert sykmelding ved 1 måned. Tid er da beregnet fra 1 måned, det vil si at 1 måned tilsvarer 2 måneder og ½ år tilsvarer 7 måneder.

b Gitt at de var på full sykmelding eller gradert sykmelding ved ½ år. Tid er da beregnet fra ½ år, slik at 1 måned tilsvarer 7 måneder og ½ år tilsvarer 1 fullt år.

## 5. Diskusjon

Vi har vist at retur fra sykmelding til arbeid ikke er en direkte og enkel prosess. Det er ikke slik at personer er sykmeldt for så å returnere til jobb, og bli der. Svært mange skifter gjentatte ganger mellom sykmeldinger og jobb. Videre kan personer skifte mellom gradert sykmelding og full sykmelding, også gjentatte ganger. Et tenkt eksempel kan være at en person debuterer med en full sykmelding og er det en periode, før personen får gradert sykmelding en periode, før igjen personen skifter tilbake til full sykmelding og til sist går tilbake til jobb.

Hvordan ble så sannsynligheten for å returnere til jobb påvirket av at denne personen var innom gradert sykmelding en viss tid?

De aller fleste debuterer (starter) i et sykefraværsløp med en full (100 %) sykmelding (se tabell 4), kanskje med mer eller mindre akutte lidelser. Hvem som etter hvert skifter til gradert sykmelding er uvisst. Det kan være en seleksjon av mer friske personer (med en viss restarbeidsevne), men det kan også være motsatt med sykere personer som har mer kroniske og langvarige lidelser. (I denne rapporten er ikke problemstillinger rundt ulike diagnoser diskutert videre). Det kan også diskuteres om gradert sykmelding i seg selv kan være et suksesskriterium, siden personene da er delvis i jobb, men i denne rapporten regnes suksess i utgangspunktet som tilbake til jobb og dermed det å ikke være på noen form for sykmelding. Et argument for dette er at gradert sykmelding ikke er ment å skulle vedvare og tilbake i jobb er således det endelige målet.

Vi vil med denne rapporten synliggjøre hvordan forløpet for retur til jobb kan være, samtidig som vi argumenterer for at dette bør og kan tas hensyn til i presentasjonen og analysene av dette. Samtidig er ikke analysene i denne rapporten ment å skulle forklare alle sider ved sykmelding og gradert sykmelding og bruken av disse. Hva effekten av å være på gradert sykemelding i seg selv er det ikke mulig å si basert på våre analyser.

Vi fant at generelt så har personer på gradert (det vil si mindre enn 100 %) sykmelding lavere sannsynlighet for å være tilbake i jobb enn personer på full (det vil si 100 %) sykmelding. Dette gjaldt både om personene startet med en gradert sykmelding, om de var gradert sykmeldt 1 måned (30 dager) senere eller om de var gradert sykmeldt ½ år (182 dager) senere.

Det er meget mulig at personer på gradert sykmelding er selektert, både når det gjelder de som starter med en gradert sykmelding og de som blir gradert sykmeldt senere, men det er ikke slik at det er friskere personer som selekteres til gradert sykmelding. Siden ulikhetene vedvarte også om vi beregner sannsynligheter gitt om de er gradert eller fullt sykmeldt ved 1 måned og ½ år, er det ikke slik at dette kan forklares ved at det er en seleksjon ved starten av en periode med sykmeldinger.

Basert på dette er det rimeligvis to mulige forklaringer, enten rekrutteres (selekteres) det personer med mer langvarige sykefravær (eventuelt lidelser) til gradert sykmelding – eller så er det slik at det faktisk er en negativ effekt av det å være på gradert sykmelding for det å returnere til jobb. En analyse innenfor ulike diagnoser kan kanskje belyse dette ytterligere, siden man da kan se om det samme mønsteret gjelder også innenfor de ulike diagnosene. Men også innenfor ulike diagnose kan det være slik at personer selekteres til gradert sykmelding hvis de har mer langvarige og mer komorbide tilstander. Analyser for ulike diagnosekategorier kan altså kanskje gi et mer nyanser bilde av dette og man kan undersøke om dette er diagnosespesifikt. Om det alternativt er slik at det faktisk er negativt å være gradert sykmeldt for å komme tilbake til jobb er vanskelig (eller umulig) å belyse basert på de tilgjengelige dataene. For å se på den faktiske effekten av å være gradert sykmeldt i forhold til fullt sykmeldt på det å returnere til jobb kreves det data av en annen kvalitet (Lie 2009).

For kjønn fant vi at kvinner har høyere sykefravær enn menn, dette er kjent. Med de utvidede analysene vi har gjennomført viser vi at det ikke er noen forskjell i intensitet ("hurtigheten") mellom menn og kvinner for å returnere tilbake til arbeid. Årsaken til forskjell i sykefravær mellom menn og kvinner er at kvinner har en høyere intensitet for å bli sykmeldt. Basert på det kan man anta at det ikke er ulikheter i det umiddelbare tilbudet (for eksempel helsetilbud) menn og kvinner får for å returnere til jobb, men at kvinner oftere blir sykmeldt. Det vil si at hvis man skal redusere sykefravær for kvinner er det altså på årsaker for å bli sykmeldt (kanskje sosiale forhold og arbeidsforhold) man bør fokusere.

Det er kjent at sykefravær øker med økende alder. Dette finner også vi. Det var imidlertid en liten forskjell i intensiteten (hyppigheten) for å bli sykmeldt for de ulike alderkategoriene vi studerte. Derimot var det en vesentlig forskjell i intensitetene for å skifte fra sykmeldt tilbake til jobb. Her fant vi en avtagende intensitet med økende alder. Det vil si at jo eldre arbeidstageren er jo lavere er intensiteten for å komme tilbake til jobb.

Dette kan for eksempel forklares ved at eldre generelt har lengre sykmeldingsepisoder enn yngre. Det kan være flere årsaker til dette, men det kan for eksempel være slik at eldre har mer langvarige lidelser enn yngre og således får lengre sykmeldinger. Dette er altså en effekt i motsatte retning av de ulikhetene vi fant mellom menn og kvinner. Som et resultat av dette bør kanskje tiltak for å redusere sykefravær for eldre fokusere på tiltak for å få eldre tilbake til arbeid (det kan blant annet si tiltak for bedring av helse).

Analysene for kalenderår viser at det ikke var en vesensforskjell i hverken intensiteter for skifte mellom jobb, gradert sykmelding og full sykmelding og derfor heller ikke i sannsynlighetene for å være i de ulike tilstandene for de to periodene 2002-2004 og 2005-2007. For den siste perioden 2008-2009 fant vi derimot at det var en høyere intensitet å bli sykmeldt og en lavere intensitet å returnere til jobb. Dette medfører at det er en lavere sannsynlighet for å være tilbake i jobb og da en høyere sannsynlighet for å være sykmeldt for denne perioden. Det er verdt å bemerke at analysene her baserer seg på de som faktisk har hatt sykmeldinger i løpet av tidsperioden 2002 og ut 2009. Det var 22.5 % av de som ble valgt ut i utvalget som ikke har sykmeldinger i det hele tatt i denne perioden. Disse som inkluderes i de generelle oversikter for andel av befolkningen som er sykmeldt, men ikke i analysene våre (<http://www.ssb.no/sykefratot/>). Selv om de statistiske analysene presentert her tar hensyn til om personer faktisk er på en av de oppgitte tilstandene (i figur 2.2) eller ikke, kan endringer i bruk av andre tiltak (det vil si bruk av rehabilitering, attføring, osv.) gjøre at det blir skjevheter over tid. Generelt er det ingenting i veien for at de relativt enkle analysene som presenteres her kan utvides til å også inkludere flere utfall. Modellen i figur 2.2 kan altså i prinsippet utvides til også inkludere flere ytelser (arbeidsavklaring, uførhet, osv.).

I regjeringens ekspertutvalg legges det til grunn at siden langvarig sykmelding gir dårlig prognose for tilbakevending til arbeidet, er det grunn til å tro at gradert sykmelding vil gi bedre prognose for tilbakevending til arbeid (Mykletun, Eriksen et al. 2010). Erfaringene med gradert tilbakevending til arbeidet er derimot splittet. Det er ikke gjennomført kontrollerte undersøkelser for å evaluere effekten av gradering. En fersk rapport fra Danmark viser en statistisk signifikant økt sannsynlighet for å komme tilbake til fullt arbeid ved det planmessige program i Danmark for en planmessig gradert retur til arbeidslivet (Hogelund, Holm et al. 2010). Det er også godt dokumentert at gradert trening er nyttig for pasienter med korsryggsmerter (Airaksinen, Brox et al. 2006). Gradert sykmelding er ikke en planmessig og systematisk gradert retur til arbeidslivet og altså ikke det samme som gradert aktivitet. Data fra Norge (Brage and Kann 2006) viser derimot en klar sammenheng mellom langt sykefravær og bruk av gradert sykmelding, slik vi også finner.

I analysene som er foretatt i denne rapporten er det kun presentert relativt enkle inndelinger av mulige årsaksfaktorer til forskjeller i sykefravær for å forenkle fortolkningen og demonstrere de analytiske perspektivene. Videre analyser for ulike diagnoser og for eksempel bruk av dialogmøter kan være aktuelle. Analysene gjort her er kun gjort for en variabel om gangen. Dette er gjort for at analysene skal framstå forståelig og vise hovedtendenser. Analysene her er videre ment å synliggjøre og fange opp prosessene for retur til arbeid. Ved blant annet å synliggjøre at øyeblikkelig *raskt tilbake til jobb fra sykmelding* (det vil si høy intensitet) ikke nødvendigvis betyr det samme som *raskt tilbake til jobb fra sykmelding og bli i jobb*, siden det å være i jobb også avhenger av hvor raskt man på ny blir sykmeldt igjen og eventuell også avgang til andre tilstander (her først og fremst gradert sykmelding). Basert på våre analyser er det også et viktig poeng å vise at personer på gradert sykmelding faktisk har en lavere sannsynlighet for å komme tilbake til og være i jobb.



## 6. Oppsummering/Konklusjon

I løpet av 8-års perioden fra 2002 og ut 2009 var det 22.5 % av de registrerte arbeidstagerne som ikke var registrert med noe sykefravær. Basert på de utvidede forløpsanalysene av de 77.5 % personene med minst en sykmelding, fant vi at personer generelt kommer raskt tilbake til jobb ved sykmelding og at cirka 70 % er tilbake i jobb 1 måned etter full sykmelding. Som regel starter ikke sykmeldingsperioder med en gradert sykmelding og personer på gradert sykmelding hadde en lavere sannsynlighet for å returnere til arbeid sammenlignet med personer på full sykmelding. Cirka 30 % av de som starter sykmeldingsperioder med en gradert sykmelding var tilbake til jobb ved 1 måned.

Vi fant at kvinner har høyere sykefravær enn menn. Dette er fordi kvinner oftere skifter fra jobb til sykmeldt (det vil si at de oftere blir sykmeldt), men når kvinner først er sykmeldt skifter de like raskt tilbake til jobb som menn (det vil si at selve varigheten av de enkelte sykmeldingsepisoder ikke er ulikt for menn og kvinner).

Sykefravær øker med alder, men det er ikke fordi eldre særlig oftere blir sykmeldt, men fordi med økende alder skifter personer seinere/sjeldnere fra sykmeldt tilbake til jobb. Det vil altså si at eldre ikke oftere blir sykmeldt enn yngre, men når de først blir sykmeldt er de sykmeldt lenger.

## Referanser

- Airaksinen, O., J. I. Brox, et al. (2006). Chapter 4 - European guidelines for the management of chronic nonspecific low back pain. *European Spine Journal*; 15: S192-S300.
- Andersen, P. K., B. Borgan, et al. (1993). *Statistical Models Based on Counting Processes*. New York: Springer-Verlag.
- Andersen, P. K. and N. Keiding (2002). Multi-state models for event history analysis. *Statistical methods in medical research*.; 11(2): 91-115.
- Andersen, P. K. and J. P. Klein (2007). Regression analysis for multistate models based on a pseudo-value approach, with applications to bone marrow transplantation studies. *Scandinavian Journal of Statistics*; 34(1): 3-16.
- Brage, S. and I. C. Kann (2006). *Fastlegers sykmeldingspraksis I: Variasjoner*: RTV-rapport 05/2006. Oslo, Rikstrygdeverket.
- Commenges, D. (1999). Multi-state models in epidemiology. *Lifetime Data Analysis*; 5(4): 315-327.
- Commenges, D. (2002). Inference for multi-state models from interval-censored data. *Statistical Methods in Medical Research*; 11(2): 167-182.
- Cox, D. R. (1972). Regression Models and Life-Tables. *Journal of the Royal Statistical Society Series B-Statistical Methodology*; 34(2): 187-&.
- Fine, J. P. and R. J. Gray (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*; 94(446): 496-509.
- Hensing, G., K. Alexanderson, et al. (1998). How to measure sickness absence? Literature review and suggestion of five basic measures. *Scandinavian journal of social medicine*.; 26(2): 133-44.
- Hogelund, J., A. Holm, et al. (2010). Does graded return-to-work improve sick-listed workers' chance of returning to regular working hours? *Journal of Health Economics*; 29(1): 158-169.
- Lie, S. A. (2009). Evaluering av tiltak for å redusere sykefravær. *Norsk Epidemiologi*; 19(2): 153-160.
- Lie, S. A., H. R. Eriksen, et al. (2008). A multi-state model for sick-leave data applied to a randomized control trial study of low back pain. *Scandinavian journal of public health*; 36(3): 279-83.
- Mykletun, A., H. R. Eriksen, et al. (2010). *Tiltak for reduksjon i sykefravær: Aktiviserings- og nærværsreform. Ekspertgrupperapport til Arbeidsdepartementet*. Oslo: Arbeidsdepartementet.
- Scheike, T. H. and M. J. Zhang (2007). Direct modelling of regression effects for transition probabilities in multistate models. *Scandinavian Journal of Statistics*; 34(1): 17-32.
- Scheike, T. H. and M. J. Zhang (2008). Flexible competing risks regression modeling and goodness-of-fit. *Lifetime Data Analysis*; 14(4): 464-483.
- Aalen, O. O. (1989). A Linear-Regression Model for the Analysis of Life Times. *Statistics in Medicine*; 8(8): 907-925.
- Aalen, O. O., Ø. Borgan, et al. (2008). *Survival and Event History Analyses: A Process Point of View*. New York: Springer-Verlag.
- Aalen, O. O. and S. Johansen (1978). An empirical transition matrix for nonhomogeneous Markov chains based on censored observations. *Scandinavian Journal of Statistics*; 5: 141-150.