



Effekt og kostnadseffektivitet av sykdomsmodifiserende legemidler ved multippel sklerose

En vurdering av to rapporter fra Folkehelseinstituttet publisert i 2019 på oppdrag fra Roche

Om Oslo Economics

Oslo Economics utreder økonomiske problemstillinger og gir råd til bedrifter, myndigheter og organisasjoner. Våre analyser kan være et beslutningsgrunnlag for myndighetene, et informasjonsgrunnlag i rettslige prosesser, eller et grunnlag for interesseorganisasjoner som ønsker å påvirke sine rammebetingelser. Vi forstår problemstillingene som oppstår i skjæringspunktet mellom marked og politikk.

Oslo Economics er et samfunnsøkonomisk rådgivningsmiljø med erfarne konsulenter med bakgrunn fra offentlig forvaltning og ulike forsknings- og analysemiljøer. Vi tilbyr innsikt og analyse basert på bransjeerfaring, sterk fagkompetanse og et omfattende nettverk av samarbeidspartnere.

Helseøkonomi og metodevurdering

Oslo Economics tilbyr økonomisk analyse av helsetjenester, teknologier og legemidler for selskaper, organisasjoner, helsetjenestetilbydere og myndigheter. Vi har kompetanse på både samfunnsøkonomiske og helseøkonomiske analyser i henhold til gjeldende veiledere, der vi identifiserer, vurderer og sammenstiller virkninger av det aktuelle tiltak.

Vi tilbyr økonomisk evaluering av helsetiltak, herunder kostnadseffektivitets-analyser og metodevurderinger. Vi har omfattende erfaring med simuleringsmodeller og bruk av real world evidence fra norske helseregistre i våre analyser.

Effekt og kostnadseffektivitet av sykdomsmodifiserende legemidler ved multippel sklerose

© Oslo Economics, 4. oktober 2019

Kontaktperson:

Erik Magnus Sæther / Sr. partner

ems@osloeconomics.no, Tel. +47 94058192

Innhold

Forord	4
Sammendrag og konklusjoner	5
1. Oslo Economics' arbeidsmetode	7
2. FHIs effektrapport (FHI 2019a)	8
3. FHIs helseøkonomiske analyse (FHI 2019b)	9
3.1 Målsetting	9
3.2 Modellstruktur	9
3.3 Sykdommens naturlige forløp i FHIs modell	10
3.4 Oslo Economics kopi-modell	10
3.5 Alvorlighet	12
3.6 Livskvalitet	12
3.7 Behandlingseffekt	12
3.8 Behandlingskostnader	13
3.9 Kostnadseffektivitet	15
3.10 Sensitivitetsanalyser	13
3.11 Budsjettvirkning	14
3.12 FHI-rapportens diskusjonsdel	14
4. Diskusjon	15
4.1 Konsekvenser for kostnadseffektivitet	15
4.2 Konklusjon	15
5. Referanser	16
Vedlegg A – Expanded Disability Status Scale (EDSS)	17
Vedlegg B – Statistiker Henrik Støvrings vurdering av FHIs nettverksmetaanalyse	18

Forord

Multipel sklerose (MS) er en svært alvorlig sykdom. De senere år har det blitt utviklet en rekke sykdomsmodifiserende legemidler som kan forsinke sykdomsutviklingen. Helseforvaltningens «Nye metoder» rekvirerte i april 2018 en fullstendig metodevurdering av disse inkludert rituksimab. Dette resulterte i to rapporter fra Folkehelseinstituttet (FHI) offentliggjort i august 2019 (FHI 2019a og FHI 2019b). Oslo Economics er av legemiddelfirmaet Roche Norge AS bedt om å gjøre en vurdering av de to rapportene. Vår vurdering presenteres i herværende rapport. Arbeidet er utført av Ivar Sønbo Kristiansen (prosjektleder), Francisco Oteiza og Erik Magnus Sæther. Statistiker Henrik Støvring har gjort en vurdering av de statistiske metoder FHI har benyttet (se vedlegg). I arbeidet med rapporten har vi utviklet en enkel Excel-modell for å simulere forløpet av MS og validere den modellen Folkehelseinstituttet har anvendt. Vår Excel-modell kan fås ved henvendelse til Oslo Economics.

Sammendrag og konklusjoner

Oslo Economics har vurdert Folkehelseinstituttets rapport om effekt av sykdomsmodifiserende legemidler (FHI 2019a; «Disease-modifying treatments for relapsing remitting multiple sclerosis, including rituximab. A health technology assessment») og den tilhørende helseøkonomirapport (FHI 2019b; «Health Economic Evaluation: Disease modifying treatments for relapsing remitting multiple sclerosis») ved såkalt «relapserende og remitterende multippel sklerose» (RRMS). FHIs analyser er utført for Nye metoder etter oppdrag fra Bestillerforum (RHF-ene) og innenfor deres tidsrammer.

Rapportene er oversiktlige, men forutsetninger og analyse er til dels lite transparente slik at kan være vanskelig å forstå og etterprøve metode og resultater. Den omfattende bruken av sladding i helseøkonomirapporteringen (FHI 2019b) bidrar i særlig grad til mangel på transparens.

I denne rapporten redegjør vurderer vi antakelser og analyser i de offentlige versjonene av rapportene. For å kunne beskrive sykdomsutviklingen i FHI-modellen, har vi utviklet en kopi-modell med input-data fra FHI-2016-rapporten og effektdata fra 2019-rapporten (FHI 2016 og FHI2019b). Vi identifiserer svakheter som beslutningstakere og lesere av FHIs rapporter bør ta hensyn til.

Nettverksmetaanalysen (effektanslag) (FHI 2019a)

- Rapporten er av god kvalitet, men metodebeskrivelsen kunne ha vært mer detaljert.
- Det er vanskelig å identifisere hvilken statistisk modell som er anvendt i nettverksmetaanalysen (NMA). Dette gjør det vanskelig å replisere analysen.
- Rapporten tar relevante forbehold for graden av evidens som det har vært mulig å inkludere i rapporten, mens den i mindre grad tar forbehold for usikkerheten i resultatene knyttet til modellvalg for NMA.
- Forskjell på behandlingseffekt mellom legemidler og mellom ulike studier kan skyldes ulikhet i pasientsammensetning (alder, kjønn, sykdomsvarighet, sykdomsgrad, behandlingsvarighet, mv). Dette confoundingproblemet er i prinsippet løst ved den metoden FHI har valgt, men da under forutsetning om såkalt transitivitet. Det skal altså ikke være interaksjon mellom pasientsammensetning og de effekter man observerer. Denne forutsetningen er antakelig ikke oppfylt i praksis.
- Rangeringen av legemidlene virker troverdig, men usikker. FHI-rapporten redegjør for denne usikkerheten.
- Evidens for rituximab er primært hentet fra en vel gjennomført svensk registerstudie basert på propensity scores for å sikre sammenlignbare behandlingsarmer. Det er usikkert hvilken betydning det har hatt for FHIs nettverksmetaanalyse at studien inngår etter oppsplitting at INF- β /GA behandlingsarmen i to armer. Denne usikkerheten burde ha fremgått tydeligere av FHIs rapport.
- Rapporten mangler analyser av hvor lenge pasientene har stått på behandling og i hvilke doser. Leseren vet derfor ikke hvilke doser og behandlingsvarigheter som har gitt de effekter man observerer i de kliniske studiene. Denne begrensningen har betydning for de påfølgende helseøkonomiske analyser.
- Rapporten mangler forløpsanalyser (time-to-event-analyser) som kan belyse tidsforløpet for sykdomsforverrelse og behandlingseffekt. Også denne type analyser har betydning for de påfølgende helseøkonomiske analyser.

Helseøkonomirapporten (kostnadseffektivitet) (FHI 2019 b)

- Folkehelseinstituttets simuleringsmodell er logisk, og strukturen er godt tilpasset sykdommen og dens varierende forløp. Beskrivelse av sykdomsutvikling ved hjelp av EDSS (expanded disability status scale) er standard i helseøkonomisk evaluering av MS.
- Modellens antagelser om livskvalitet på ulike sykdomsstadier (EDSS-nivåer) ved behandlingsstart virker rimelige. EDSS-0 betyr at pasienten er symptomfri mens EDSS-9 betyr at han/hun er helt sengeliggende (se vedlegg A).

- Modellen er så vidt vi kjenner til ikke validert mot norske eller utenlandske data om MS-forløpet selv om den er publisert internasjonalt.
- Modellen beregner at etter 20 års sykdom kan 20% av pasientene sitte oppe noen timer daglig mens de ellers er sengeliggende (EDSS-8), 12% er helt sengeliggende (EDSS-9) mens 7% er døde (EDSS-10). Dette betyr et langt mer alvorlig sykdomsforløp enn hva norske og utenlandske studier tyder på eller hva norske klinikere synes er rimelig.
- Rapportens omtale av behandlingsvarighet er knapp, men synes å innebære at pasientene, med unntak for to legemidler, behandles med det samme sykdomsmodifiserende legemiddel i 20 år. Denne antagelsen stemmer dårlig med virkeligheten og mangler faktagrunnlag. I praksis bytter noen pasienter legemiddel på grunn av bivirkninger eller manglende effekt, men dette fanges ikke opp i modellen og er heller ikke diskutert.
- Rapportens omtale av behandlingseffekter er knapp, men analysene synes å være basert på en antagelse om effekt på alle EDSS-nivåer. Forskningslitteraturen tyder derimot på at sykdomsmodifiserende behandling bare virker på de tidligere stadier av RRMS og ikke i det hele tatt på SPMS.
- Helsetjenestekostnadene i ulike stadier av MS er basert på en norsk studie av pasientenes egne utgifter, men fanger ikke opp de betydelige kostnader helsetjenesten har, særlig i de senere stadier av sykdommen. FHIs analyse konkluderer at 20 års behandling av MS uten sykdomsmodifiserende legemidler bare koster kroner 718.875 (4% kalkulasjonsrente) hvilket virker urealistisk lavt fordi pasientene får kostnadskrevende diagnostikk og behandling i alle faser av sykdommen. FHI har i en tidligere rapport anslått den samme kostnaden til 5,9 millioner kroner. Selv om bare de færreste pasienter er helt sengeliggende etter 20 år, vil hver pasient innlagt på institusjon alene ha kostnader på minst 1,5 millioner kroner per år.
- De avvik som er nevnt ovenfor kan ha stor betydning for beregning av både kostnader og helsegevinster. Når modellen overvurderer sykdomsforverring, blir kostnadene overvurdert og helsegevinstene feilvurdert. Når modellen overvurderer behandlingsvarighet, blir både behandlingskostnader og helsegevinster overvurdert. Når kostnadene ved ikke-legemiddel-behandling av MS er undervurdert, blir kostnadsbesparelsene ved effektiv behandling undervurdert.
- Rapporten presenterer ingen sensitivitets- eller scenariorapporter som kan belyse hvilke inputs i modellen som driver resultatene. Den presenterer probabilistiske sensitivitetsanalyser, men disse er sladdet.
- I motsetning til Legemiddelverkets metodevurderinger, er resultater med godkjent maksimalpris for legemidlene ikke vist (sladdet). Denne mangelen, sammen med sladding av hovedresultater, gjør det umulig for leseren å vurdere og tolke kostnadstall og kostnadseffektivitet, selv om man analyserer med godkjent maksimalpris.

Den helseøkonomiske rapporten følger hverken norske eller internasjonale retningslinjer for rapportering av helseøkonomiske analyser (SLV 2018, Husereau 2013). Omtale av metode er kortfattet og enkelte input-data er overhodet ikke vist i tekst- eller tabellform. Enkelte av våre vurderinger må derfor sees i lys av at vi kan ha misforstått de metoder FHI har anvendt.

Oppsummert synes den helseøkonomiske modellen å passe dårlig med sykdommens naturlige forløp, og det blir da høyst usikkert om modellens resultater blir rimelige både hva angår kostnader og helsegevinster. Behandlingsområdet er krevende å modellere, og vi er innforstått med at FHI hadde knappe tidsfrister for sine analyser. Vi mener likevel at rapporten, i sin nåværende form, er uegnet til å treffe beslutninger om prioritering eller valg av behandling. Da modellens logikk og struktur er godt tilpasset problemstillingen, kan den justeres slik at den gir realistisk sykdomsprogresjon, behandlingskostnader og helsegevinster.

1. Oslo Economics' arbeidsmetode

Vi har gjennomgått Folkehelseinstituttets hovedrapport med beregning av legemiddeffekter (FHI 2019a) og helseøkonomiske rapport (FHI 2019b). Vi har gjennomgått aktuelle referanser i disse samt søkt etter annen relevant litteratur. Referanselisten viser hvilke artikler og rapporter vi har vurdert. Da den helseøkonomiske rapporten ikke viser hvordan modellen simulerer gradvis forverring av multipel sklerose (MS) over tid, har vi utviklet en modell-kopi i Excel på basis av de input-data FHI har brukt i sin modell. Dette er gjort for å sammenligne FHI-modellens simuleringer av sykdomsforverring med hva man vet fra forskningslitteraturen. En ekstern statistiker har vurdert nettverksmetaanalysen (se vedlegg).

2. FHIs effektrapport (FHI 2019a)

I vedlegg nedenfor presenterer vi en grundig gjennomgang av nettverksmetaanalysen (FHI 2019a). I tillegg drøfter vi her noen begrensninger ved effektrapporten der disse har betydning for den helseøkonomiske analysen.

Legemiddelkostnadene avhenger av hvor lenge pasientene i praksis bruker legemiddelet og i hvilke doser. Informasjon om faktiske doser, medikamentetterlevelse (adherence) og behandlingsavslutning blir presentert i rapporter og artikler fra kliniske studier og registerstudier, om enn i varierende grad. Ingen slike data blir presentert i FHIs effektrapport. Denne type data har nødvendigvis betydning for tallfesting så vel av kostnader som behandlingseffekter. I mangel av denne type data synes den helseøkonomiske rapporten å være basert på antagelsen om at pasientene, med to unntak, bruker de aktuelle sykdomsmodifiserende legemidler sammenhengende i 20 år. I praksis er behandlingsvarighet langt kortere. Bivirkninger og manglende effekt gjør at legemiddelbehandling blir avsluttet. Behandlingsbytte er aktuelt når bedre eller mer kostnadseffektive behandlinger kommer på markedet. Slike endringer over tid kan modelleres i helseøkonomiske modeller hvilket har vært gjort blant annet ved bruk av biologiske legemidler for inflammatorisk leddsykdom (leddgikt, Bechterews sykdom mv).

FHIs helseøkonomiske analyse bruker kvalitetsjusterte leveår som effektmål hvilket er et fornuftig og vanlig valg. Det betyr at analysen kan fange opp hvor raskt behandlingen gir effekt hvilket har en klar betydning for pasientene. FHIs nettverksmetaanalyse er basert på metoder der eventuell variasjon i tidsforløp av effekter ikke fanges opp. Slik effekter krever forløpsanalyser (time-to-event-analyser) hvilket ikke er anvendt.

3. FHIs helseøkonomiske analyse (FHI 2019b)

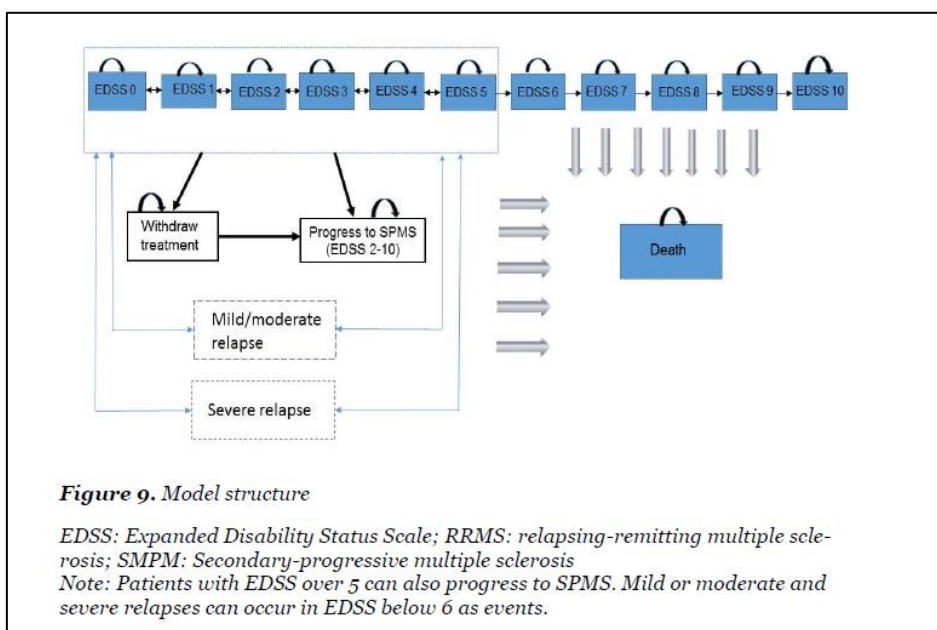
Den helseøkonomiske rapporten «Health Economic Evaluation: Disease modifying treatments for relapsing remitting multiple sclerosis» (FHI 2019b) skal gi Beslutningsforum og andre beslutningstakere et grunnlag for prioritering og behandlingsvalg.

3.1 Målsetting

Rapporten har som mål å vurdere kostnadseffektivitet for 10 ulike legemiddelbehandlinger inklusive rituksimab som ikke har RRMS som godkjent indikasjon. Rapporten er et eksempel på en fullstendig metodevurdering der man ikke bare sammenligner to, men mange behandlinger. Dette gir et bedre grunnlag for prioriteringer, men gjør også oppgaven mer krevende.

3.2 Modellstruktur

Modellen beskriver forløpet av såkalt relapserende remitterende multippel sklerose (RRMS) fra tidlig fase av sykdommen til langt fremskredet stadium der pasientene kan være helt sengeliggende med økt dødelighet. Sykdomsforløpet beskrives med såkalt EDSS-score der EDSS-0 betyr symptomfrihet med gradvis økning mot EDSS-9, der pasienten er helt sengeliggende, til EDSS-10 som betyr død av MS. Pasientene kan oppleve forverring for hvert år som går, men kan også forbli på samme EDSS-nivå i to eller flere år eller oppleve forbedring. Modellen fanger opp legemiddelkostnader og andre behandlingskostnader i hvert enkelt av de 20 år modellen følger pasientene. De sistnevnte øker med økende EDSS-score. Modellen antar at noen pasienter går over i en sekundær progressiv fase (SPMS) der sykdomsprogresjonen går raskere. Den fanger også opp akutte forverrelser som krever tilleggsbehandling og representerer tilleggskostnader. Modellstrukturen virker logisk og synes å passe svært godt med det kliniske forløp av sykdommen dersom bare input-tallene er korrekte.



Figur 1. Prinsippskisse for den helseøkonomiske modell. Kilde: FHI 2016b.

Modellen beregner 20 års forløp av MS hvilket virker som en rimelig avveining mellom datatilgjengelighet og behovet for langsiktige analyser. Den antar en rimelig fordeling av pasientene blant de laveste EDSS-nivåer ved start av simuleringen. Modellen fanger opp legemiddelkostnader og andre helsetjenestekostnader, leveår og kvalitetsjusterte leveår. Den beregner samlede kostnader og samlede kvalitetsjusterte leveår (QALYs) over 20 år uten behandling eller for hver av de aktuelle legemidler. Til slutt uttrykkes kostnadseffektiviteten for hver behandling som forholdet mellom merkostnad og mereffekt sammenlignet med ingen behandling eller med de øvrige legemidler.

3.3 Sykdommens naturlige forløp i FHIs modell

Forverrelse av MS uttrykkes i modellen ved årlig sannsynlighet for å gå fra et sykdomsstadium (EDSS-nivå) til det neste, typisk i en såkalt Markovmodell. Det finnes ingen norske studier der man har fulgt større grupper av MS-pasienter over flere tiår, hverken før eller etter ca. 1995 da sykdomsmodifiserende legemidler kom på markedet. I mangel av norske data henter FHIs modell forløpsdata fra en studie i Ontario, Canada, der man fulgte en stor pasientgruppe over mer enn 30 år (Weinschenker 1989; Kremenchutzky 2006). Disse observasjonsdataene har vært brukt i flere utenlandske studier og synes å være et anerkjent datagrunnlag.

Markovmodellens input-data for sykdommens naturlige forløp (forverrelse i EDSS-status) er presentert i en helseøkonomisk rapport fra 2016 (FHI 2016). Overgang fra et EDSS-nivå til et annet er beskrevet med rater (hendelser per observasjonsår), tabellene 16-19). Markovmodeller er basert overgangssannsynligheter og ikke rater, men vi antar at ratene er omregnet til overgangssannsynligheter slik korrekt er.

Sannsynligheten både for forverrelse i RRMS og overgang til SPMS virker til dels nokså høye med overgangsrater opp til 0,485 (overgangssannsynlighet 0,384). Pasientenes forventede antall kvalitetsjusterte leveår (QALYs) fra 30-årsalder til død (ca 45 leveår) er bare 11,0 QALYs. Man får derfor et intuitivt inntrykk av sykdommen forverres ganske raskt tross at det også er lagt til grunn 8,3% sannsynlighet for forbedring i visse pasientgrupper. Dersom FHIs modell overvurderer sykdomsforverringen ved MS, kan det få stor betydning for beregning både av helsegevinster og kostnader. Pasientene kommer tidligere over i svært kostnadskrevede stadier, men høy dødelighet vil trekke i motsatt retning, mens helsegevinstene kan bli større. Rapporten presenterer ingen validering av modellen eller sammenligning med andre publiserte modeller. FHIs 2016-analyse er riktignok publisert som artikkel i et vitenskapelig tidsskrift (Hamidi 2018), men artikkelen gir ikke reviewerne grunnlag for validering. Vi utviklet derfor en enkel kopi av FHIs modell for å kunne vurdere om sykdomsutviklingen virker rimelig i forhold til hva man ville vente ut fra forskningslitteraturen.

3.4 Oslo Economics kopi-modell

For å kunne beskrive sykdomsutviklingen i FHI-modellen, har vi utviklet en kopi-modell med input-data fra FHI-2016-rapporten og effektdata fra 2019-rapporten (FHI 2016 og FHI2019b).

Det er uklart om FHIs resultater er basert på sannsynligheter (hvilket er korrekt) eller på de rater (hendelse per år) rapportene angir. Vi har derfor gjort analyser med begge tolkninger, men antar i det følgende at FHI har brukt sannsynligheter. Vi har brukt følgende datatyper i vår kopi-modell:

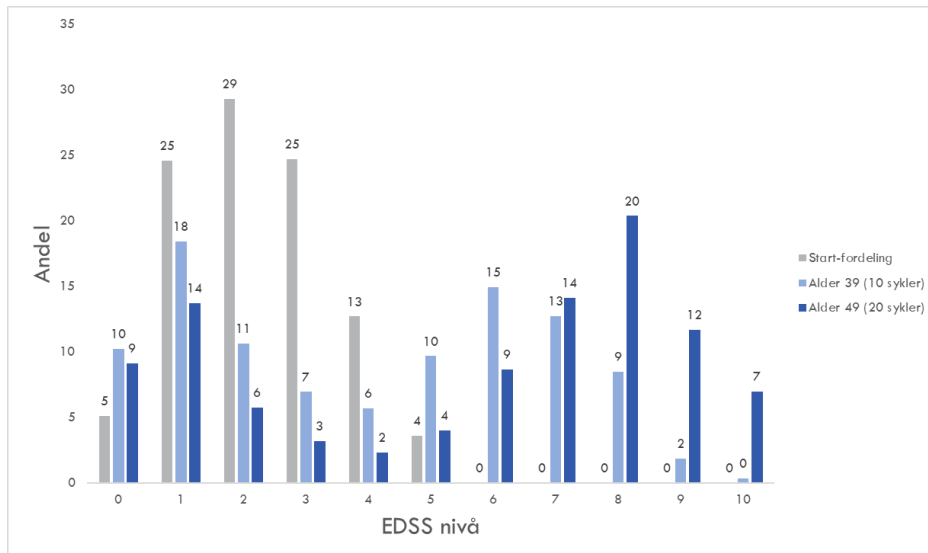
- Årlig sannsynlighet for forverrelse av RRMS (forverrelse i EDSS-nivå)
- Årlig sannsynlighet for overgang til progressiv sykdom (SPMS)
- Årlig sannsynlighet for forverrelse av SPMS (forverrelse i EDSS-nivå)
- Årlig sannsynligheten for forbedring av RRMS (forbedring i EDSS-nivå)
- Effekt av legemiddelbehandling (relativ risiko for forverring, mindre enn 1,0)

Vi har, i tråd med FHIs metode, anvendt samme sannsynligheter for forverring i alle 20 år. Dette er neppe realistisk, men ikke drøftet i FHIs rapport (FHI 2019b).

Når det gjelder behandlingseffekt, antok vi at behandlingen påvirket progresjon i RRMS på alle EDSS-nivåer. Vi antok at legemidler ikke påvirker overgang til progressiv sykdom (SPMS) eller forverrelse når man har fått progressiv sykdom. Vi antok at pasientene, med unntak for alemtuzumab og kladribin, fikk sykdomsmodifiserende behandling i 20 år og at den relative effekten var konstant over alle 20 år (se nedenfor om kostnader). Da vi primært var interessert i sykdomsprogresjon, la vi ikke inn attakker («relapses») i modellen fordi disse påvirker kostnader og kvalitetsjusterte leveår, men ikke sykdomsprogresjon utover sannsynligheten for forverrelse som allerede ligger i modellen. Vår kopi-modell er utviklet i Excel og kan fås ved henvendelse til Oslo Economics.

Figur 2 viser EDSS-fordelingen ved simuleringsstart og etter 10 og 20 år. Modellen tilsier at etter 20 år er 20% av pasientene i EDSS-nivå 8 (kan sitte oppe, men er mest sengeliggende), 12% i EDSS-nivå 9 (helt sengeliggende) mens 7% er i EDSS-tilstand 10 (død av MS). Modellen tilsier altså en særdeles dyster utvikling av sykdommen uten behandling. Vi har derfor sammenlignet disse resultatene med publisert litteratur.

Figur 2. Fordeling (prosentandel) etter sykdomsgrad (EDSS-nivå) ved start og etter 10 og 20 år uten behandling



I en oppfølging av kanadiske MS-pasienter fant man at ca 55% av pasientene hadde gått fra RRMS til SPMS etter 20 år (Scalfari 2014). Vår kopi-modell viste 63%. Den samme studien viste at i størrelsesorden 25-30% (avlest på en graf) av pasienter med progresjon til SPMS var progrediert til EDSS-8 eller verre. Kopi-modellen viste 57%. En annen studie basert på de samme data viste at 2-3% (avlest på en graf) av pasienter med RRMS er døde etter 20 år (Kremenchutzky 1999). En norsk studie viser (avlest på graf) ca 5% døde (uansett årsak) etter 20 år (Lundet 2017). Kopi-modellen tilsier 7% døde av MS. I alle sammenligningene viser kopi-modellen mer alvorlig sykdomsforløp enn de data modellen er bygget på.

I en norsk studie fulgte man 50 pasienter med RRMS og progressiv MS over 10 år (Jacobsen 2014). Pasientene de fikk MS-diagnosen i perioden 1998-2000 og fikk altså i liten grad sykdomsmodifiserende behandling. Etter 10 år var gledeligvis 50% av pasientene uten sykdomsprogresjon. FHI-modellen tilsier at mindre, trolig lang mindre, enn 50% er uforandret etter 10 år.

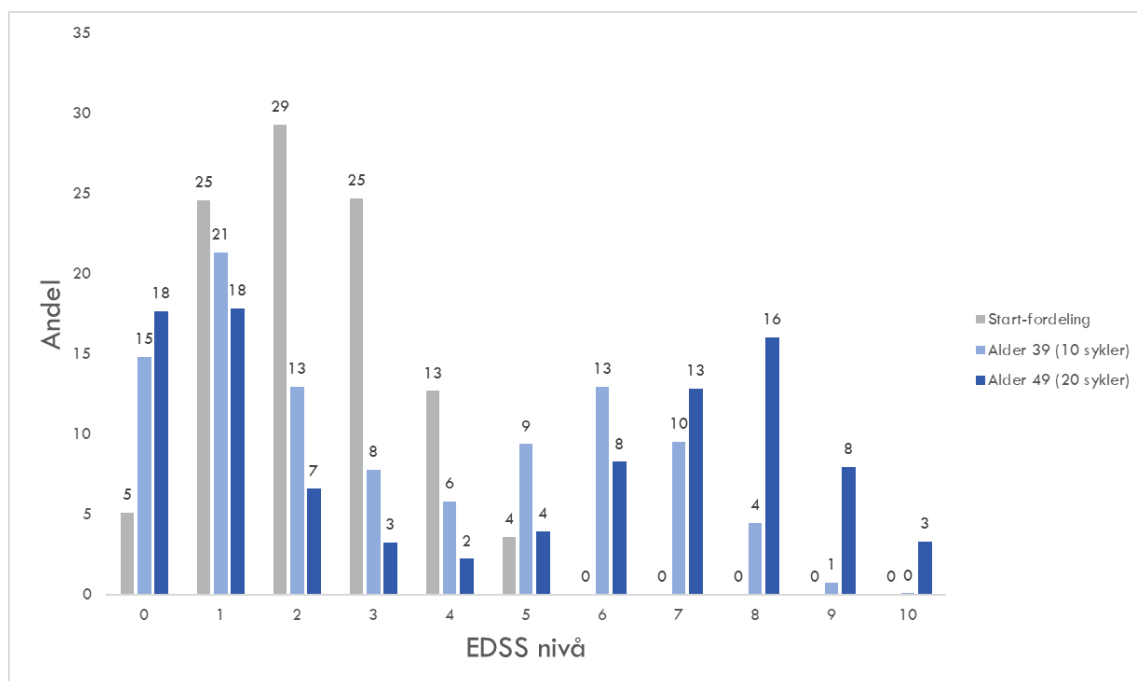
Den kanadiske studien som den norske helseøkonomimodellen bygger på, inkluderte 1099 MS pasienter i perioden 1972-84. Det er rimelig å anta at medisinsk behandling og rehabilitering av MS-pasienter har blitt bedre fra 1970-årene selv uten sykdomsmodifiserende behandling.

Det ser altså ut til at FHI-modellen simulerer et forløp av RRMS som er mer alvorlig enn de registerdata den er bygget. Inntrykket bekreftes av norske forskningsartikler og samtale med norske nevrologer. Når modellen overvurderer sykdomsprogresjonen, betyr det at den undervurderer hvor lenge pasientene lever og den overvurderer sannsynligvis helsegevinstene ved behandling. Når modellen overvurderer sykdomsforverrelse trekker det i retning av å overvurdere kostnadene fordi disse er høyest i de mest alvorlige stadier av sykdommen. Vi tror modellen også overvurderer dødeligheten, og dette vil trekke i motsatt retning på kostnadene.

Figur 3 viser EDSS-fordeling når pasientene settes på det legemiddelet som hadde størst effekt i FHIs effektrapport (okrelizumab)(FHI 2019a). Mens andelen pasienter i EDSS-gruppene 8, 9 og 10 er 39% uten behandling, er den 27% med behandling. Som ventet blir sykdomsprogresjonen altså betydelig mindre. Vi mangler data for å kunne vurdere om endringen er rimelig. Hvis vi har forstått FHI-rapporten korrekt (FHI 2019b), antar FHI at sykdomsmodifiserende behandling blir gitt på alle EDSS-nivåer og både ved RRMS og SPMS. I praksis blir slik behandling mest brukt ved lave EDSS-nivåer (Chalmer 2018). Få sykdomsmodifiserende legemidler er aktuelle ved SPMS. I FHIs rapport fra 2016 er det riktignok anført at pasienter med SPMS og

EDSS-nivå på 6 eller høyere ikke får behandling (FHI 2016). Det er mulig at 2018 rapporten også bruker denne antagelse selv om det ikke er nevnt.

Figur 3. Fordeling (prosentandel) etter sykdomsgrad (EDSS-nivå) ved behandlingsstart og etter 10 og 20 år med behandling (okrelizumab, relativ progresjonsrisiko 0,53)



3.5 Alvorlighet

FHI anslår at 30 år gamle pasienter med RRMS bare vil oppnå 11,0 kvalitetsjusterte leveår (QALYs) i sin resterende livstid. Dette utgjør et tap av 32,1 QALYS i forhold til de 43,1 QALYs befolkningen ellers kan regne med i 30-årsalder. Fordi FHIs modell overvurderer sykdomsforverrelsen, er QALY-tapet ved MS overdrevet, men det er likevel ingen tvil om at MS er en svært alvorlig sykdom.

3.6 Livskvalitet

Modellen uttrykker effekt som kvalitetsjusterte leveår (QALYs). De anslag FHI benytter virker meget rimelige.

3.7 Behandlingseffekt

FHIs helseøkonomiske analyse baseres på «annualized relapse rate» (attakk-rate) og «risk of disability progression» (FHI 2019a). FHIs metaanalyse viser også «change in disability (EDSS score) from baseline (FHI 2019a), men disse estimatene brukes ikke i den helseøkonomiske modell uten at dette kommenteres. FHI antar i den helseøkonomiske analyse at sannsynligheten for attakk og for forverring er den samme i alle 20 år etter behandlingsstart. Dette er neppe realistisk. En annen Markov-modell varierer progresjon etter alder og antar at pasienten kan gjøre mer enn ett EDSS-steg opp eller ned per år (Palace 2014).

FHIs meta-analyse presenterer effekt av de ulike legemidler på dokumentert sykdomsforverrelse etter 3-6 måneder (varierende definisjon i ulike studier)(FHI 2019b). Disse effekt-tallene brukes i den helseøkonomiske analysen som effekt på 12-måneders sannsynlighet for EDSS-forverrelse hvilket trolig er forskjellig fra 3-6 måneders forverrelse. Meta-analysen presenterer også tall for EDSS-endring, men de brukes ikke i den helseøkonomiske analysen. Ingen av disse metodeproblemer diskuteres i den helseøkonomiske rapporten.

3.8 Behandlingskostnader

Legemiddelkostnader er vanligvis en dominerende kostnadsfaktor i legemiddeløkonomiske analyser. FHIs rapport er neppe noe unntak selv om nesten alle kostnadsresultatene er sladdet. Behandlingsvarighet er avgjørende for legemiddelkostnadene (og effektene). Rapporten anfører: "Most drugs will be administered every year and have the same drug cost each year under treatment". Dette kan neppe tolkes annerledes enn at FHI antar at pasientene blir stående på det samme sykdomsmodifiserende legemiddel i 20 år med mindre de dør i perioden. Et unntak er kladribin og alemtuzumab der behandlingen er langt mer kortvarig, men hvor en viss andel av pasientene behandles på ny hvert år. I FHIs analyse fra 2016 (FHI 2016) har man antatt et visst bortfall fra behandling de to første år, men dette er ikke nevnt i årets rapport (FHI 2019b).

Generelt er det vanlig at pasienter avslutter legemiddelbehandling på grunn av bivirkninger, manglende effekt eller fordi man vil forsøke en annen behandling, og behandlingsvarighet kan være overraskende kort. Det er all grunn til å tro at dette også gjelder behandling av RRMS fordi behandlingene kan ha alvorlige bivirkninger. En svensk registerstudie viste at ca 10% av pasienter som starter på rituksimab, har sluttet etter 2 år mens over 40% på interferon hadde sluttet med interferonbehandling (Spelman 2018). Andelen for interferonbehandling var over 40%. FHI modellanalyse og antagelser om behandlingsvarighet, avviker derfor, hvis vi har forstått dem korrekt, betydelig fra virkeligheten. I den grad behandlingen avsluttes før 20-årsperioden er over, blir legemiddelkostnadene lavere, andre behandlingskostnader kan bli høyere på grunn av sykdomsforverrelse og helsegevinstene blir mindre.

FHIs analyser fanger opp både legemiddelkostnader og andre behandlingskostnader. FHI-rapporten viser imidlertid ingen tabeller som kan vise hvilke kostnader som inngår i analysene. FHI har gjort analyser både med godkjent maksimalpris for legemidler og de faktiske priser etter rabatter, men resultat-tallene er sladdet. Behandlingskostnader utover legemidler er basert på en norsk studie der pasienter rapporterte sine personlige kostnader knyttet til MS (Svendsen 2018), men ikke vist i rapporten. I forskningsartikkelen fra studien er kostnader vist etter EDSS-nivå, men studien fanger bare opp pasientens utgifter. Når pasientenes samlede egenandeler tross alt er beskjedne i forhold til de samlede helsetjenestekostnader, betyr det at analysene grovt undervurderer behandlingskostnadene.

Den norske kostnadsundersøkelsen viser at pasienter på EDSS-nivå 9 har utgifter på 11.875 euro per år (Svendsen 2018). Disse pasientene er sengeliggende og bor enten i sykehjem eller bor hjemme med hjemmesykepleie, hjemmetjenester og eventuelt brukerstyrt personlig pasient. Når man tar i betraktning at ett års opphold i sykehjem koster ca 1,4 millioner kroner¹ og hjemmebehandling vil ligge på samme nivå, blir FHIs anslag for helsetjenestekostnadene ved MS-behandling altfor lave. FHIs rapport anslår at 20 års helsetjenester ved MS koster 718.885 kroner ved 4% kalkulasjonsrente. Selv om de største kostnadene kommer sent i forløpet der diskonteringen får størst betydning, er det liten tvil om at FHIs anslag er altfor lavt.

I samsvar med norske retningslinjer for legemiddeløkonomiske analyser (SLV 2018) baseres legemiddelkostnadene på dagens priser. Konkurransen fra andre produkter og tap av anbud gir vanligvis betydelig prisfall. På den annen side kan arbeidskraftkostnader tenkes å øke over tid på grunn av økende etterspørsel etter arbeidskraft. Slike prisendringer kan få betydelig innvirkning på livstidskostnadene ved behandling. Retningslinjene sier også at produksjonsgevinster ved at pasientene kan holdes lenger yrkesaktive heller ikke skal hensyntas hvilket har betydning, noe som kan ha særlig stor betydning for sykdommer som MS. Dette til forskjell fra norsk praksis for samfunnsøkonomiske analyser på andre områder (samferdselsprosjekter mv).

3.9 Sensitivitetsanalyser

FHI-rapporten presenterer fem probabilistiske sensitivitetsanalyser. Alle er imidlertid sladdet slik at leserne ikke kan se resultatene. Det er i helseøkonomiske analyser vanlig å belyse hvordan endringer i enkeltvariable eller grupper av variable (for eksempel sannsynlighet for EDSS-forverrelse) påvirker resultatene. Når slike sensitivitetsanalyser ikke er vist i rapporten, kan leseren derfor ikke danne seg noen formening om hva som driver modellresultatene og hvilken type usikkerhet som er viktigst. I særdeleshet vil leseren være interessert i informasjon om hvilke overgangssannsynligheter som er viktigst for resultatene.

¹ Norske sykehjem kostet i 2018 i alt 55.104 millioner kroner og hadde 39.572 plasser (foreløpig tall fra SSB). Dette tilsvarer 1.392.500 kroner per plass per år

3.10 Budsjettvirkning

FHI angir at man har benyttet anslag fra Legemiddelverket for rituksimab og okrelizumab for å beregne budsjettvirkningene av disse to legemidlene. Når det gjelder metode for budsjettanalyser, presenterer FHI dette over tre linjer, men disse er sladdet. Så vidt vi kan se er det ikke presentert noen tall for budsjettvirkninger i rapporten. Leseren kan altså vurdere hverken metode for budsjettanalysene eller se resultatene.

Vi vil generelt si at anslag for budsjettvirkning må ventes å være lavere enn behandlingsretningslinjer for behandling skulle tilsi. Forklaringen er at pasientene av ulike grunner slutter med behandlingen tidligere enn anbefalt eller bruker lavere dose enn forskrevet. Dessuten faller prisene som følge av priskonkurranse og tap av patent. Oslo Economics har nylig gjort en gjennomgang av Legemiddelverkets metodevurderinger av PD1/PDL1-behandling. Denne viste at de faktiske, samlede kostnader for medikamentgruppen var mindre enn halvparten av Legemiddelverkets anslag, selv etter at budsjettanalysene var justert for antatt rabattnivå (<https://www.dagensmedisin.no/contentassets/ca12168fdd754b2e8051c8186fc69272/3.-ivar-sonbo-kristiansenimmunoterapinorgearendalsukaisk13aug2019.pdf>).

3.11 FHI-rapportens diskusjonsdel

Rapportens diskusjonsdel drøfter enkelte metodeproblemer, særlig beregning av kostnader ved legemiddelinfusjoner. Den drøfter imidlertid ikke de sentrale problemene knyttet til sannsynligheten for sykdomsprogresjon og om modellen er i samsvar med hva man mener å vite om sykdomsforløpet. Den drøfter heller ikke hva kliniske studier viser om behandlingsvarighet eller hvorvidt det er realistisk at pasientene står på samme sykdomsmodifiserende behandling i 20 år. Medline har 436 treff på nøkkelordene cost-effectiveness og multiple sclerosis. Selv om de fleste av disse er trolig mindre relevante for den norske problemstillingen, ville det ha vært nyttig å sammenligne utenlandske anslag for QALY-gevinster med FHIs anslag.

4. Diskusjon

Vår gjennomgang av FHIs helseøkonomiske rapport tyder på en rekke og potensielt alvorlige svakheter knyttet til simulering av sykdomsforløp, behandlingsvarighet og helsetjenestekostnader. Disse feil og svakheter betyr at både kostnads- og effektanslag i FHI-rapporten kan avvike betydelig fra de faktiske verdier. Et gjennomgående problem er at metode og input-data ikke er presentert i samsvar med retningslinjer for rapportering av helseøkonomiske analyser (SLV 2018, Husereau 2013). Når metodene er såpass knapt fremstilt, kan vi ikke utelukke at vi har misforstått parametriseringen av modellen og tar forbehold om dette.

4.1 Konsekvenser for kostnadseffektivitet

Etter vår oppfatning er FHIs helseøkonomiske analyse bygget på en modell som egner seg godt til å belyse kostnadseffektiviteten ved MS-behandling. Den har imidlertid flere alvorlige svakheter som drøftet ovenfor. Disse kan medføre betydelig feil i beregningen av så vel kostnader som behandlingseffekter.

På kostnadssiden er antagelsen om 20 års behandling uten skifte til annen sykdomsmodifiserende behandling nokså urealistisk. Dette betyr at både legemiddelkostnadene og helsegevinstene blir overvurdert. Mer realistiske antagelser om behandlingsvarighet vil bringe ned de totale behandlingstkostnader, men samtidig vil behandlingseffekten bli lavere. De øvrige helsetjenestekostnader betydelig undervurdert fordi de stort sett bare omfatter pasientens egne utgifter, ikke helsetjenestens kostnader.

Når modellen tilsier at en betydelig del av pasientene etter 20 år både med og særlig uten behandling er rullestolbundet, sengeliggende eller døde, betyr at den overvurderer sykdomsprogresjonen. Det betyr trolig at modellen også overvurderer behandlingseffekter. Hva de samlede effekter av alle disse svakheter betyr for kostnader og effekter av det enkelte legemiddel og deres relative kostnadseffektivitet, kan man ikke vurdere før det er gjort en modellanalyse med mer realistiske forutsetninger.

4.2 Kostnadseffektivitet

Vurdering av kostnadseffektivitet forutsetter at man har beregnet samlede langtidskostnader (legemidler og andre kostnader) og helsegevinst målt som kvalitetsjusterte leveår. De fleste kostnadstall i FHI-rapporten er sladdet (FHI 2019b). FHIs modell tilsier at pasientene, sammenlignet med ingen behandling, kan vinne mellom 0,2774 (Glatiramer 40mg) og 1,2004 (okrelizumab) QALYs ved sykdomsmodifiserende behandling. Fremtidige QALY-gevinster er justert med 4% kalkulasjonsrente, og gevinstene er således meget betydelige. Validiteten av tallene er imidlertid usikker fordi modellen ikke passer med den sykdomsprogresjon man kan vente i Norge. I motsetning til Legemiddelverket presenterer ikke FHI samlede langtidskostnader basert på godkjente, offentlige maksimalpriser for legemidlene. Rapporten anslår at 20 års behandling av RRMS uten bruk av sykdomsmodifiserende legemidler kan ventes å koste 718.885 kroner diskontert 4% årlig. Dette anslaget er åpenbart for lavt og sterkt i strid med FHIs tidligere anslag.

Da FHI ikke presenterer tall for langtidskostnader sammen med langtidseffekter (QALY-gevinster), er det ikke mulig å gjøre en fullstendig vurdering av kostnadseffektiviteten. I omtalen av de probabilistiske sensitivitetsanalysene, som er sladdet, angis enkelte rangeringer. En fullstendig vurdering av kostnadseffektiviteten ved de ulike sykdomsmodifiserende legemidlene krever imidlertid at man har korrekte tall for både langtidskostnader og langtidseffekter for alle de aktuelle behandlingene. Leseren av rapporten har således ingen mulighet til en fullstendig vurdering av kostnadseffektivitet selv om grunnlagstallene skulle være korrekte.

4.3 Konklusjon

Oppsummert synes den helseøkonomiske modellen, slik den er parametrisert, å passe dårlig med sykdommens naturlige forløp, og det blir da høyst usikkert om modellens resultater blir rimelige både hva angår kostnader og helsegevinster. Behandlingsområdet er krevende å modellere, og vi er innforstått med at FHI hadde knappe tidsfrister for sine analyser. Vi mener likevel at rapporten, i sin nåværende form, er uegnet til å treffe beslutninger om prioritering eller valg av behandling. Da modellens logikk og struktur er godt tilpasset problemstillingen, kan den justeres slik at den gir realistisk sykdomsprogresjon, behandlingstkostnader og helsegevinster.

5. Referanser

- Chalmer TA et al. Early versus later treatment start in multiple sclerosis. *Eur J Neurol* 2018; 25: 1262.
- FH 2016. Couto E, Hamidi V, Ringerike T, Odgaard-Jensen J, Harboe I, Klemp M. NIPH Systematic Reviews. In: *Medicines Used for Multiple Sclerosis - A Health Technology Assessment*. Oslo: Folkehelseinstituttet, 2016.
- FH 2019b. Hagen G, Højslev LU, Fretheim A, Hamidi V. *Health Economic Evaluation: Disease modifying treatments for relapsing remitting multiple sclerosis*. Oslo: Folkehelseinstituttet, 2019.
- FH 2019a. Tjelle TE, Rose CJ, Ohm IK, Pike E, Håheim LL, Bidonde J, Fretheim A, Juvet LK. *Disease-modifying treatments for relapsing remitting multiple sclerosis, including rituximab. A health technology assessment*. Oslo: Folkehelseinstituttet, 2019.
- Hamidi V et al. A Multiple Treatment Comparison of Eleven Disease-Modifying Drugs Used for Multiple Sclerosis. *Journal of clinical medicine research* 2018; 10: 88-105.
- Husereau et al. Consolidated health economic evaluation reporting standards (CHEERS). *Value Health*. 2013; 16(2): 231-250.
- Jacobsen C et al. Brain atrophy and disability progression in multiple sclerosis patients. *J Neurol Neurosurg Psychiatry* 2014; 85: 1109-15.
- Kremenutzky-M et al. The natural history of MS: a geographically based study 7: Progressive-relapsing and relapsing-progressive multiple sclerosis – a re-evaluation. *Brain* 1999; 122: 1941-9.
- Kremenutzky-M et al. The natural history of MS: a geographically based study 9: Observations of the progressive phase of the disease. *Brain* 2006; 129: 584-94.
- Lunde-HMB et al. Survival and cause of death in multiple sclerosis. *J Neurol Neurosurg Psychiatry* 2017; 88: 621-25.
- Palace J et al. UK multiple sclerosis risk-sharing scheme: a new natural history dataset and an improved Markov model. *BMJ Open* 2014; 4: e004073.
- Scalfari A et al. Onset of secondary progressive phase and long-term evolution of multiple sclerosis. *J Neurol Neurosurg Psychiatry* 2014; 85: 67-75.
- SLV 2018. *Retningslinjer for dokumentasjonsgrunnlag for hurtig metodevurdering av legemidler*. Oslo: Statens legemiddelverk, 2018.
- Svendsen B, Grytten N, Bo L, Aarseth H, Smedal T, Myhr KM. The economic impact of multiple sclerosis to the patients and their families in Norway. *Europ J Health Econ. HEPAC : health economics in prevention and care* 2018;19(9):1243-57.
- Spelman T et al. Comparative effectiveness of rituximab relative to IFN-beta or glatiramer acetate. *Multiple Sclerosis Journal* 2018; 24: 1087-95.
- Weinschenker BG et al. The natural history of multiple sclerosis: A geographically based study. *Brain* 1989; 112: 133-46.

Vedlegg A – Expanded Disability Status Scale (EDSS)

Score	Description
1.0-1.5	No disability, minimal signs in one FS
2	Minimal disability in one FS
2.5	Mild disability in one FS or minimal disability in two FS
3	Moderate disability in one FS, or mild disability in three or four FS. No impairment to walking
3.5	Moderate disability in one FS and more than minimal disability in several others. No impairment to walking
4	Significant disability but self-sufficient and up and about some 12 hours a day. Able to walk without aid or rest for 500m
4.5	Significant disability but up and about much of the day, able to work a full day, may otherwise have some limitation of full activity or require minimal assistance. Able to walk without aid or rest for 300m
5	Disability severe enough to impair full daily activities and ability to work a full day without special provisions. Able to walk without aid or rest for 200m
5.5	Disability severe enough to preclude full daily activities. Able to walk without aid or rest for 100m
6	Requires a walking aid – cane, crutch, etc. – to walk about 100m with or without resting
6.5	Requires two walking aids – pair of canes, crutches, etc. – to walk about 20m without resting
7	Unable to walk beyond approximately 5m even with aid. Essentially restricted to wheelchair; though wheels self in standard wheelchair and transfers alone. Up and about in wheelchair some 12 hours a day
7.5	Unable to take more than a few steps. Restricted to wheelchair and may need aid in transferring. Can wheel self but cannot carry on in standard wheelchair for a full day and may require a motorised wheelchair
8	Essentially restricted to bed or chair or pushed in wheelchair. May be out of bed itself much of the day. Retains many self-care functions. Generally has effective use of arms
8.5	Essentially restricted to bed much of day. Has some effective use of arms retains some self-care functions
9	Confined to bed. Can still communicate and eat
9.5	Confined to bed and totally dependent. Unable to communicate effectively or eat/swallow

Vedlegg B – Statistiker Henrik Støvring's vurdering av FHIs nettverksmetaanalyse

Evaluering af Folkehelseinstituttets rapport om effekt av sykdomsmodifiserende legemidler – behandlingseffekt af rituximab

Dato: 3. oktober 2019

Henrik Støvring
Stats4all
Runevej 53
8210 Aarhus V
Danmark
CVR 33272359

Baggrund og formål

Multipel Sclerose (MS) er en alvorlig sykdom med et langt forløb karakteriseret ved gentagne relaps og remissionsperioder, hvor sværhedsgraden af relaps forsterkes, når patienten har gjennomgået mange tidligere tilbakefald. MS diagnosticeres typisk i 30-års alderen og følges så af et 20-30 årigt sykdomsforløb. Patienter kan typisk forvente at miste 5-10 års levetid på grund af sykdommen. I det lange sykdomsforløb vil de fleste patienter gjennomgå behandlingsperioder med forskjellige medikamenter, hvor valget af et medikament på et givent tidspunkt vil afhænge af, hvordan patienten har reagert på de medikamenter, patienten tidligere har modtaget. Behandlingsforløb er derfor ofte komplekse og kan vanskelig sammenlignes direkte mellom forskjellige patienter/patienttyper.

FHI har i sin rapport lavet en network meta-analyse (NMA) af de medikamenter, som er relevante i en norsk kontekst (FH 2019a). Da kun få randomiserede studier er gjennomført med direkte sammenligninger af aktive, moderne behandlinger inddrages også observationelle studier i analysen. Medikamenterne sammenlignes primært på fire aspekter: Relaps, progression, alvorlige bivirkninger (SAE) og withdrawal. Disse fire aspekter rapporteres i Radar plots, som giver et overblik over et medikaments samlede virkningsprofil (figur 26, side 57). Generelt bør polygoner med størst areal foretrekkes, om end ikke alle fire aspekter nødvendigvis bør tillægges samme vøgt. Rituximab 500/1000 mg har sammen med alemtuzumab 12 mg, natalizumab 300 mg og ocrelizumab 300 mg nogenlunde samme størrelse polygon. De tre sidste synes at have bedre effekt i forhold til relaps og progression, mens rituximab ser ut til at have en fordelagtig bivirkningsprofil. Rapporten vurderer dog, at den spesifikke rangering er usikker og at særlig graden af evidens for effekterne af rituximab er lav, da den primære evidens for rituximab stammer fra et observationelt studie.

Formålet med dette notat er at vurdere denne konklusjon.

Vurdering af NMA

FHI beskriver tre overordnede tilgange til analyse, som repræsenterer tre forskjellige versjoner af NMA. Den analytiske tilgang er basert på, at hver behandlingsarm i hvert studie indgår med et estimat og en standard error (SE, mål for præcision). Hvor der ikke har været opplyst en SE direkte, har FHI selv beregnet den ud fra konfidensintervaller (CI) ved bruk af standardformler for sammenhængen mellom bredden af CI og SE. Hvor utfaldene i trials har været rater (fx annualised relapse rate) er estimerne blevet modelleret som normalfordelte på log-skala (FH 2019a, side 80), hvilket er standard jf Poisson-baserte analyser af incidens- og mortalitetsrater.

For at beskrive den anvendte NMA med random effects, som foretrekkes af FHI (den tredje model, Table 11, FH 2019a, side 83) kan vi anvende flg notation: Vi lader det j te inkluderede studies estimat for relapsraten i behandlingsarm i med behandling k være givet ved r_{ijk} med tilhørende SE ($SE(r_{ijk})$). Modellen, som beskrives af FHI, kan da skrives som (FH 2019a, side 82 nederst):

$$\log(r_{ijk}) = b_0 + b_k + a \cdot I(NRS) + \gamma_j + \eta_{[NRS:j]} + \epsilon$$

I denne model-ligning er b_0 log-raten for reference-behandling (Placebo – FH 2019a, side 84), b_k er kontrasten på log-rate skala mellem behandling k og reference, a er en effekt som ikke-randomiserede studier gennemsnitligt afviger fra randomiserede studier med. Alle disse effekter er systematiske, dvs. de udtrykker gennemsnitlige log-rater (b_0) eller gennemsnitlige forskelle mellem log-rater (b_k og a). γ_j er en random effekt for studie j , dvs en effekt som deles mellem armene i studie j . Tilsvarende er $\eta_{[NRS:j]}$ en random effekt, som repræsenterer variation imellem randomiserede studier og ikke-randomiserede studier. Beskrivelsen af denne effekt er ikke helt klar ("...and between randomised and non-randomised evidence" (FH 2019a, side 82)) da man ved læsning af Efthimiou et al (96), som FHI rapporten henviser til har, får indtryk af, at denne effekt repræsenterer en random effekt for hhv randomiserede studier og ikke randomiserede studier (dvs sådan som det er skrevet i ovenstående modelligning). ϵ er den uforklarede variation i NMA modellen, hvilket vil sige, hvordan det enkelte studies estimat afviger fra det forventede ud fra modellen. Da forekomsten af events generelt synes lav i de inkluderede studier vil en analyse baseret på rater (eller risiko) være tilnærmelsesvis sammenlignelige med resultater baseret på HR.

Ovenstående model sikrer i princippet, at effekten fra hvert studies interne sammenligning bevares, da der er en studiespecifik random-effekt, som modellerer, hvordan et givent studie i begge arme afviger fra de andre studier. Der er derfor ikke behov for at korrigerer for patientsammensætning og andre kovariater, som varierer i mellem studier, da disse så at sige opfanges af den studiespecifikke random effect. Dette forudsætter dog (som NMA generelt forudsætter), at der ikke er interaktion mellem patientmix og effekten der ses i det enkelte studie (transitivitet). Denne antagelse kan for eksempel være problematisk, hvis varighed af behandling og dermed effekt af behandling varierer mellem studierne hvilket må antages at være tilfældet her. FHI-rapporten har ikke analyseret medikament-overlevelse (adherence) som formodes at variere mellem medikamenter og studier. Behandlingseffekten vil afhænge sygdommens progression, hvilket kan sætte spørgsmålstegn ved antagelsen om, at der ikke interaktion mellem patientsammensætning og medikamenteffekt. I princippet vil den af FHI anvendt model sikre, at der korrigeres for en systematisk forskel mellem randomiserede og ikke-randomiserede studier, samt at der kan være forskellig grad af heterogenitet mellem studier i de to kategorier. I protokollen omtales modellen som en variant af en generalized linear model (GLM), hvor estimationen sker med vægte baseret på SE (Tjelle, n.d.).

Selv om ovenstående repræsenterer et kvalificeret bud på den model, som FHI har anvendt, ville det være ønskeligt, hvis FHI eksplicit havde angivet den anvendte model. FHI anerkender selv, at modellen er kompleks og at de har haft behov for at søge ekstern bistand til implementering af modellen i en angivet statistikpakke (FH 2019a, side 83). Det bidrager heller ikke til afklaring, at modellen beskrives som en "...frequentist arm-wise version of the class of models proposed by Efthimiou et al. (96) to combine randomised and non-randomised evidence in an NMA." (FH 2019a, side 83) I Efthimiou referencen, der henvises til, optræder flere versioner af modeller for denne analysetilgang og det er ikke klart, præcist hvilken af disse, som søges emuleret af FHIs version (Efthimiou et al. 2017). I protokollen for FHIs analyse henvises primært til andre artikler, nemlig (Rücker 2012; Schwarzer, Carpenter, and Rücker 2015), jævnfør (Tjelle, n.d.). Hverken i protokol eller rapport fremgår den præcise matematiske modelligning, og uden denne vurderes det vanskeligt at replicere analysen nøjagtigt. Det er dog en styrke ved rapporten, at der i appendiks rapporteres effektestimater fra de forskellige modeller, som FHI har anvendt i analysen, og der ses generelt konsistens mellem effektestimater fra forskellige modeller.

Det er en styrke ved undersøgelsen, at rapporten dokumenterer afvigelser fra den præ-specificerede protokol. Afvigelse synes primært at skyldes nødvendigheden af at foretage pragmatiske modelleringsvalg i lyset af den sparsomme underliggende evidens, der har været tilgængelig for inklusion i FHIs analyse. En afvigelse fra protokollen er indhentning af data direkte fra forfatterne af et svensk baseret studie (Spelman et al. 2018), hvor INF- β /GA gruppens relapse events og risikotid splittes op på de to behandlinger. Det er vanskeligt at vurdere konsekvensen af dette. På den ene side sikrer den, at sammenligningen med andre studier ikke sløres af en heterogent sammensat aktiv komparatorgruppe. På den anden side bryder det den balancering mellem behandlingsarmene, som internt i studiet er opnået ved at anvende propensity scores til at matche patienter. Det er vanskeligt at sige i hvilken retning, dette vil påvirke estimaterne i NMA. Generelt ses en tendens til, at NMA finder effekter med bredere sikkerhedsintervaller end i de individuelle studier, især hvor graden af evidens er lav. Dette er forventeligt, da sammenligning af studier fra forskellige settings er mere usikre end interne sammenligninger i et enkelt studie. Dette kan formodentlig forklare forskellen mellem fx Rituximab vs INF- β /GA hvad disability progression angår. Spelman 2018 finder i den direkte sammenligning en HR på 0.86 (0.52; 1.43) (Figure 2, Spelman et al. 2018), mens FHI angiver estimater i størrelsesorden 0.56 til 0.66, men med væsentligt bredere sikkerhedsintervaller, fx (0.21; 2.02 for Rituximab vs INF- β) (Figure 9, side 33, FH 2019a). Der er

således ikke umiddelbart modstrid mellem de forskellige estimater, da de alle er upræcise, ikke mindst baseret på NMA. At sikkerhedsintervaller er bredere baseret på NMA er forventeligt, da de også afspejler heterogenitet mellem studier – effekt af af INF- β relativt til placebo vil således kunne forventes at variere mellem forskellige studier, men estimeres som en effekt.

Vurdering af evidensen baseret på observationelle studier, særligt med hensyn til rituximab

Der er kun udført et randomiseret fase II studie af effekten af rituximab ved behandling af multipel sclerose, og det er derfor vigtigt at vurdere validiteten af de observationelle, ikke-randomiserede studier. Her er studiet af Spelman et al. 2018 centralt.

Samlet set vurderer FHI evidensen for rituximab som lav, fordi der kun er udført et randomiseret forsøg med sammenligning med Placebo, hvor designet giver høj risiko for bias (FH 2019a, side 21). At det randomiserede studie har en høj risiko for bias forekommer at være en rimelig vurdering med det lave antal patienter inkluderet i studiet og med valget af en ikke-aktiv komparator behandling (placebo). Uden yderligere randomiserede studier er evidensen derfor i sagens natur begrænset til at komme fra observationelle studier. Observationelle studier kan give pålidelig og troværdig evidens, men det forudsætter, at de har justeret og helst elimineret de relevante kilder til bias. Det er derfor vigtigt at overveje, hvilke kilder til bias, der må forventes at være til stede i observationelle studier af medikamenters effekt på MS.

Bias i observationelle studier af medikamenters effekt på MS

De bias-kilder, som forekommer at være særligt relevante i observationelle studier af behandling for MS, er:

- **Confounding by disease severity**

Det må forventes, at svært syge patienter tilbydes mere aggressiv behandling med større risiko for bivirkninger end patienter med en mildere grad af sygdom. Behandling, som er mere virksom og derfor vælges som "second line treatment", kan i naive sammenligninger nemt fremstå som en behandling, der tolereres dårligt af patienten og har lav effekt, hvis den sammenlignes direkte med en "first line treatment". Dette kan imødegås ved at sammenligne patientgrupper, som har samme fordeling af sværhedsgrad af sygdommen (den såkaldte *exchangeability* antagelse).

- **Tolerans og effekt af tidligere behandling**

At en patient tidligere har oplevet svigt af en behandling eller manglende tolerans for en behandling kan have indflydelse på valg af det følgende medikament. Dette kan give skævhed i sammenligningen mellem to medikamenter, hvis det ene medikament oftere gives til patienter, som ikke havde effekt af behandling eller ikke kunne tolerere bivirkningerne. Denne bias imødegås ved at sammenligne patientgrupper, som ideelt set har haft samme forudgående behandlingsforløb.

- **Sygdomsvarighed**

Denne bias kan vise sig i to retninger. På den ene side vil patienter, som har haft sygdommen i lang tid typisk have en mere fremskreden sygdom, det vil sige sygdom i mere svær grad. Samtidigt vil patienter, som har haft MS i lang tid, typisk være patienter, som enten har en mindre aggressivt progredierende sygdom eller er mere modstandsdygtige overfor sygdommen. Denne bias optræder, når der anvendes prævalente patientkohorter i observationelle designs, hvilket vil sige, at patienter ikke inkluderes ved første diagnose af sygdommen. I stedet inkluderes de patienter som har overlevet indtil et inklusionstidspunkt, som typisk er en kortere eller længere tidsperiode, hvor der er indsamlet registerdata for patientgruppen. Denne bias imødegås ved at sammenligne patienter med samme varighed af sygdommen ved inklusion.

For at kunne sammenligne en effekt estimeret i et observationelt studie med et tilsvarende randomiseret studie er det generelt en forudsætning, at der i det observationelle studie er anvendt et *new user design* med en relevant komparator (Lund, Richardson, and Stürmer 2015; Yoshida, Solomon, and Kim 2015). Hvis der ikke anvendes et *new user design* vil tolerans af behandling skævvride den tilsyneladende behandlingseffekt, så et medikament med dårlig tolerans fremstår som en behandling med større effekt end det i virkeligheden har – patienter som tåler medikamentet i længere tid kan være mere robuste og have større modstandskraft også over for selve sygdommen (en variant af "healthy worker survivor effect").

Vurdering af Spelman et al. 2018

I vurderingen af rituximab benytter FHI primært et observationelt studie spel (Spelman et al. 2018). Studiet finder en betydelig effekt af rituximab sammenlignet med INF- β og glatiramer acetate (GA), for så vidt angår relapsraten, der estimeres til at være 87% lavere med Rituximab sammenlignet med INF- β /GA. For *three month confirmed disability progression* var der ikke en signifikant forskel (HR 0.86 (0.52; 1.43)).

I fraværet af randomisering er analysen baseret på *propensity scores* (PS). PS estimerer sandsynligheden for, at en given patient ved baseline vil modtage rituximab blandt alle som modtager enten rituximab eller INF- β /GA. Estimationen baseres på patient karakteristika tilgængelige ved baseline. Efterfølgende matches patienter, som faktisk fik rituximab, med patienter, som fik INF- β /GA, så de matchede personer har tilnærmelsesvis samme PS. Dermed sammenlignes patienter med tilnærmelsesvis ens sandsynlighed for at modtage rituximab, hvilket efterligner et randomiseret forsøg, hvor alle patienter typisk har eksakt samme chance for at få en given behandling. Forskellen er, at PS kun skaber balancering ud fra observerede karakteristika, mens et randomiseret studie balancerer både observerede og uobserverede karakteristika.

Studiet af Spelman et al. (2018) adresserer eksplicit de bias-kilder, som er gennemgået ovenfor, og artiklen viser i Table 2, at PS tilgangen har balanceret patienternes køn, alder, sværhedsgrad af sygdom (EDSS), sygdomsvarighed, antal af tidligere behandlinger og forekomst af relapse i op til to år før inklusion i studiet. Som FHI vurderer (FH 2019a, side 21) er der derfor umiddelbart lav risiko for bias i studiets estimer af effekterne af rituximab.

De primære begrænsninger ved Spelman et al. studiet er beskrivelsen af den inkluderede studiepopulation og definitionen af, hvornår en patient er persistent med sin behandling, alternativt afbryder sin behandling (discontinuation). Det svenske MS register indeholder ifølge artiklen information om 80% af de svenske MS patienter, men da indrapportering åbenbart er frivilligt for den behandlende læge er det vanskeligt at vurdere, hvilken selektion af patientgruppen dette indebærer. Det kan ikke udelukkes, at patienter og læger som ikke indgår i registeret adskiller sig systematisk fra dem som indgår. Der er ikke dog umiddelbart grund til at tro, at selektionen skulle være knyttet til, hvilken behandling patienterne modtager. Hvad angår persistens/afbrydelse af behandling er der ikke angivet om dette defineres ud fra nogle kriterier for medicinbrug eller om det registreres af den behandlende læge. Hvis det sker ud fra kriterier, er disse ikke oplyst. Det er en styrke ved studiet at patienter i begge grupper har skullet modtage behandling i mindst tre måneder med index-medikamentet for at kunne indgå.

Det er en styrke ved studiet af Spelman et al (2018), at der er gennemført en egentlig bias-analyse, som finder at der skal være markant uobserveret confounding, for at dette vil kunne forklare de observerede associationer. Det er dog vigtigt at huske, at confounding selv i mindre omfang vil kunne forklare noget af de observerede associationer. Der er med andre ord en risiko for at effekten er overvurderet, selv om det ikke forekommer plausibelt, at den alene skyldes uobserveret confounding.

Sammenfatning

Evidensen for rituximab til behandling af MS er sparsom og består hovedsageligt af et mindre, randomiseret studie med betydelig risiko for bias, samt et observationelt studie, som viser betydelig effekt i en svensk kohorte af MS patienter (Spelman et al. 2018), hvad angår funktionsevne (EDSS). Studiet vurderes af FHI til at have lav risiko for bias. Denne vurdering forekommer rimelig, idet artiklen anvender en analyse baseret på propensity scores, der har formået at balancere observerede patientkarakteristika, og at artiklen indeholder en eksplicit bias-analyse for at vurdere, hvorvidt resultaterne kan skyldes uobserveret confounding. Ud fra rutinemæssigt anvendte Cochrane-kriterier for validiteten af evidens, nedgraderes artiklen imidlertid til at have lav sikkerhed for evidens ("we judged the certainty of evidence...to be low", (FH 2019a, side 25), da den ikke bygger på et randomiseret studie. Dette forekommer umiddelbart at afvige fra FHIs egen vurdering af, at studiet har lav risiko for bias. Randomisering er generelt det bedste værn mod uobserveret bias, men hvis resultaterne i et observationelt studie ikke skyldes bias, er det vanskeligt at se, hvorfor studiets validitet skulle være lav.

Selv om den valgte tilgang til Network Meta Analyse overordnet set virker velvalgt, er det vanskeligt entydigt at afgøre, hvilken betydning den endelige udformning af modellen har haft på resultaterne. Ingen af de anvendte NMA modeller skrives eksplicit op, og der henvises i rapporten til privat korrespondance i udformningen af den endelige model. FHI rapporten beskriver evidensen som sparsom og den udtrykker lav tillid til nogle af resultaterne. Selv om det overordnede resultat angående rangering af de forskellige medikamenter fremstår troværdigt, gør usikkerhederne i analysen det vanskeligt at give entydige behandlingsrekommendationer.

- FHIs rapport præsenterer generelt et stort og grundigt arbejde af god kvalitet, om end metodebeskrivelsen kunne være mere detaljeret.
- Det er vanskeligt at identificere den præcise model, som er anvendt i den udførte NMA, hvorfor det vil være vanskeligt at replicere analysen.

- Rapporten tager relevante forbehold for graden af evidens, det har været muligt at inkludere i rapporten, mens den i mindre grad tager forbehold for usikkerheden i resultaterne, som skyldes modelvalg for NMA.
- Rangeringen af medikamenterne forekommer troværdig, om end den er usikker. FHI rapporten gør opmærksom på denne usikkerhed.
- Evidens for rituximab er primært hentet fra et velgennemført svensk register studie baseret på propensity scores for at sikre sammenlignelige behandlingsarme. Det er usikkert, hvilken betydning det har haft for FHIs NMA, at studiet indgår efter opsplitning af INF- β /GA behandlingsarmen i to arme. Denne usikkerhed kunne have fremgået tydeligere af FHIs rapport.

Referencer

- Efthimiou, Orestis, Dimitris Mavridis, Thomas P. A. Debray, Myrto Samara, Mark Belger, George C. M. Siontis, Stefan Leucht, and Georgia Salanti. 2017. "Combining Randomized and Non-Randomized Evidence in Network Meta-Analysis." *Statistics in Medicine* 36 (8): 1210–26. <https://doi.org/10.1002/sim.7223>.
- FH 2019a. Tjelle, Torunn E, Christopher James Rose, Ingrid Kristine Ohm, Eva Pike, Lise Lund Håheim, Julia Bidonde, Atle Fretheim, and Lene Kristine Juvet. 2019. "Disease-Modifying Treatments for Relapsing Remitting Multiple Sclerosis, Including Rituximab. A Health Technology Assessment. - NIPH." Oslo, Norway.
- Lund, Jennifer L, David B Richardson, and Til Stürmer. 2015. "The Active Comparator, New User Study Design in Pharmacoepidemiology: Historical Foundations and Contemporary Application." *Current Epidemiology Reports* 2 (4): 221–28. <https://doi.org/10.1007/s40471-015-0053-5>.
- Rücker, Gerta. 2012. "Network Meta-Analysis, Electrical Networks and Graph Theory." *Research Synthesis Methods* 3 (4): 312–24. <https://doi.org/10.1002/jrsm.1058>.
- Schwarzer, Guido, James R. Carpenter, and Gerta Rücker. 2015. *Meta-Analysis with R*. Use R! Cham: Springer International Publishing. <https://doi.org/10.1007/978-3-319-21416-0>.
- Spelman, Tim, Thomas Frisell, Fredrik Piehl, and Jan Hillert. 2018. "Comparative Effectiveness of Rituximab Relative to IFN- β or Glatiramer Acetate in Relapsing-Remitting MS from the Swedish MS Registry." *Multiple Sclerosis Journal* 24 (8): 1087–95. <https://doi.org/10.1177/1352458517713668>.
- Tjelle, Torunn E. n.d. "Project Plan: Health Technology Assessment of Medicines Used for Multiple Sclerosis. Part I: RRMS." Oslo, Norway.
- Yoshida, Kazuki, Daniel H Solomon, and Seoyoung C Kim. 2015. "Active-Comparator Design and New-User Design in Observational Studies." *Nature Reviews. Rheumatology* 11 (7): 437–41. <https://doi.org/10.1038/nrrheum.2015.30>.

oslo**economics**

www.osloeconomics.no

post@osloeconomics.no
Tel: +47 21 99 28 00
Fax: +47 96 63 00 90

Besøksadresse:
Kronprinsesse Märthas plass 1
0160 Oslo

Postadresse:
Postboks 1562 Vika
0118 Oslo