

Modeller för att värdera, betala och finansiera ATMP - En samhällsekonomisk analys

Mats Bergman
Jonas Björnerstedt
Johan Stennek

18 april 2021

1	INLEDNING	4
2	ATMP – VAD ÄR ANNORLUNDA?	5
3	UTGÅNGSPUNKTER FÖR VÄRDERING AV LÄKEMEDEL OCH INSTITUTIONELLA FÖRHÅLLANDEN	9
3.1	INSTITUTIONELLA FÖRHÅLLANDEN	9
3.2	SAMHÄLSEKONOMISK KOSTNADS-INTÄKTSANALYS	11
3.3	VÄRDERING AV LIV OCH HÄLSA	12
3.4	SJUKVÅRDENS BUDGET	13
3.5	ICER, NET BENEFIT OCH MARGINALNYTTAN AV INKOMST	15
3.6	VÄRDEBASERAD PRISSÄTTNING	16
3.7	PRISSÄTTNING OCH KONKURRENS	17
3.8	VAL MELLAN ALTERNATIVA BEHANDLINGAR	18
3.9	PATIENTENS EGENSKAPER (INDIKATIONSBASERAD VÄRDERING)	20
4	BERÄKNING AV KOSTNAD PER QALY FÖR ATMP	22
4.1	ALLMÄNNA UTGÅNGSPUNKTER	23
4.1.1	<i>Väntevärdesriktig värdering</i>	23
4.1.2	<i>Oersättlighet och osäkerhet</i>	24
4.1.3	<i>Diskontering</i>	24
4.2	OSÄKERHET OM LÄKEMEDLETS EFFEKTER	25
4.3	FÖRVÄNTAD PRISNEDGÅNG FÖR JÄMFÖRELSEALTERNATIVET	26
4.4	KOSTNADER ÖVER TIDEN	27
4.5	HOPP OCH ORO	28
5	INLÅSNINGSKOSTNADER OCH BEHANDLINGSNEUTRALITET	30
5.1	BEHANDLINGSNEUTRALITET	30
5.2	VÄRDERING AV ATMP ÖVER TIDEN	31
5.3	VÄRDET AV ATT VÄNTA	33
5.4	KOSTNAD AV ATT VÄNTA	34
5.5	KOSTNAD FÖR INKREMENTELLT BÄTTRE BEHANDLINGAR	35
5.6	OPTIONSVRÄDE OCH ATT VÄNTA	36
6	KONKURRENS OCH INCITAMENT FÖR ATT INVESTERA I FOU	38
6.1	TERAPEUTISK KONKURRENS	39
6.1.1	<i>Metoder för att stimulera till terapeutisk konkurrens med ATMP</i>	39
6.2	INVESTERINGSINCITAMENT OCH KONKURRENSNEUTRALITET	40
6.2.1	<i>Investeringsincitament</i>	41
7	BETALNINGSMODELLER	43
7.1	BETALNINGSMODELLER – EN ÖVERSIKT	44
7.2	JURIDISKA FÖRUTSÄTTNINGAR FÖR BETALNINGSMODELLER FÖR ATMP	47
7.3	BETALNINGSMODELLER VID ASYMMETRISK INFORMATION. VAD KAN VI LÄRA AV TEORIN FÖR <i>MECHANISM DESIGN</i> ?	48
7.4	BEHANDLINGSNEUTRALITET OCH INFORMATION	50
7.5	BETALNINGAR BETINGADE PÅ UTFALL	50
7.6	KOSTNADER OCH SVÅRIGHETER FÖRKNIPPADE MED UTFALLSBASERAD PRISSÄTTNING	52
8	ANSVARSFÖRDELNINGEN MELLAN STAT OCH REGIONER	55
8.1	RISKFÖRDELNING	55
8.2	EFTERFRÅGAN (PATIENTBEHOV)	56
8.3	ENGÅNGSKOSTNADER FÖR ATMP	57
8.4	LÅNEMÖJLIGHETER	58
8.5	FORSKNINGSINCITAMENT	59
8.6	ADMINISTRATIV KOMPETENS	59
8.7	SUBSTITUT	59
9	REFERENSER	60

APPENDIX A. ICER, NET BENEFIT OCH MARGINALNYTTAN AV INKOMST	62
APPENDIX B. VÄRDERING AV ATMP	63
B.1. OSÄKERHET	64
B.2. DET KONTRAFAKTISKA ALTERNATIVET	65
B.3. MAXIMAL OCH FÖRVÄNTAD ERSÄTTNING	66
APPENDIX C. VÄRDET AV ATT VÄNTA	66
APPENDIX D. INVESTERINGSINCITAMENT	67
D.1. MINSKAD KONKURRENS	67
D.2 KONKURRENS OCH INVESTERINGSINCITAMENT	68
APPENDIX E. BETALNINGAR BETINGADE PÅ INFORMATION	68

1 Inledning

Syftet med denna rapport är att belysa ett antal problem relaterade till värdering, prissättning och betalning av ATMP utifrån ett nationalekonomiskt perspektiv. Vi har inte ambitionen att ge en uttömmande översikt av ATMP eller ens att beskriva de grundläggande medicinska egenskaperna, utan vårt fokus är de aspekter som vi bedömer som viktiga för att förstå hur ett effektivt regelverk för prissättning och betalning ska kunna utformas. Vi kommer också att beröra utmaningar som har uppmärksammats kring prissättning av och betalningsmodeller för ATMP. Eftersom problemområdet är stort är rapporten inte heltäckande.

Diskontering är ett komplicerat problemområde inom nationalekonomin. Särskilt för diskontering över lång tid saknas det en konsensus inom ämnet om både utgångspunkter och slutsatser i termer av vilken faktisk diskonteringsränta som ska användas. Trots detta krävs ett ställningstagande för att analysera utfall över lång tid. I vår analys gör vi det genom att presentera två rapporter. I en annan fristående underlagsrapport till uppdraget analyserar Johan Stennek diskontering och tidspreferenser utifrån grundprinciper. I denna huvudrapport utgår vi istället från diskontering enligt TLV:s praxis, för att fokusera på frågor kring värdering av ATMP.

2 ATMP – vad är annorlunda?

ATMP är en förkortning för Advanced Therapy Medicinal Products. För en utförlig beskrivning, se EMA:s hemsida.¹ De tre huvudtyperna av ATMP, samt en fjärde kategori bestående av kombinationer av dessa, beskrivs av Läkemedelsindustriföreningen på följande sätt:²

Somatiska cellterapi innehåller celler som bearbetats väsentligt så att deras egenskaper, funktion eller struktur har ändrats och att cellerna därefter ges till patient för att behandla, förebygga eller diagnosticera en sjukdom genom sin effekt.

Genterapier definieras av att de innehåller arvs massa (DNA, RNA eller annan nukleinsyrasekvens) och att de ges i syfte att reglera, reparera, ersätta, addera eller ta bort en gensekvens som orsakar sjukdom eller skada.

Vävnadstekniska produkter definieras av att de innehåller bearbetad vävnad eller celler för nybildning, reparation eller ersättning av human vävnad.

Kombinationsläkemedel som innehåller avancerade terapiläkemedel innehåller även medicintekniska produkter (t.ex. matrix eller stödkonstruktioner).

Ur hälso- och sjukvårdsekonomisk synpunkt är följande egenskaper viktiga:

1. Många ATMP behandlar ovanliga sjukdomar samtidigt som utvecklingskostnaderna är höga. Detta ger höga fasta kostnader per patient.
2. Många ATMP är individuellt anpassade, exempelvis med utgångspunkt från patientens egna celler, vilket innebär förhållandevis höga rörliga kostnader per patient.
3. ATMP är i de flesta fall engångsbehandlingar som förväntas ha positiva effekter i årtal, decennier eller livet ut. Osäkerheten om effekten och hur länge den består är dock i många fall stor.
4. Vissa ATMP ersätter ett traditionellt läkemedel som hade behövt tas livet ut; andra ATMP ersätter traditionella läkemedel eller annan typ av behandling som ges under en relativt kort period, t ex vid cancer.
5. ATMP kan vara svåra att kopiera, vilket innebär att förutsättningarna för konkurrens är sämre än för traditionella läkemedel.
6. ATMP administreras inom vården, dvs är ej förskrivningsläkemedel.
7. I vissa fall behandlar ATMP sjukdomar för vilka ingen alternativ behandling finns, eller där alternativ behandling är påtagligt sämre i olika avseenden.
8. Hittills finns väldigt få ATMP kommersiellt tillgängliga, men utvecklingen går fort och många nya ATMP förväntas nå marknaden inom några år.

Konsekvensen av ovanstående är att ATMP har potential att tillföra samhället stora värden i form av bättre hälsa och ökad överlevnad, vilket givetvis är positivt. Emellertid medför också de ovan nämnda egenskaperna en del problem som framför allt har att göra med de höga

¹ <https://www.ema.europa.eu/en/human-regulatory/marketing-authorisation/advanced-therapies/advanced-therapy-classification>

² <https://www.lif.se/nyheter/ny-arbetsgrupp-for-avancerade-terapilakemedel/>

kostnaderna per behandling samt att stor osäkerhet råder om hälsoeffekter långt fram i tiden.

I vissa fall medför detta stora nettoökningar av kostnaderna för sjukvården. I andra fall kan ATMP medföra totalt sett lägre vård- och läkemedelskostnader – men också att stora engångskostnader faller ut tidigt, medan besparingarna sprids ut, potentiellt över patientens hela återstående livstid.

Detta innebär i sin tur att sjukvården riskerar att utsättas för mycket höga kostnader under de år då många ATMP lanseras och först används för stora patientgrupper. En sådan kostnadspuckel kan vara svår för regionerna att hantera. För sjukdomar som är geografiskt koncentrerade kan enskilda regioner drabbas av särskilt höga kostnader.

Möjligen kan det uppfattas som ett problem att patienter som behandlats med ATMP flyttar mellan regioner. I någon mening tillfaller nyttan den region till vilken en patient flyttar, medan den region där patienten tidigare bodde har fått stå för hela kostnaden för ATMP. Vi menar att detta är ett mindre problem i det svenska systemet, även om det kan vara ett problem i andra länder, där sjukvården är försäkringsfinansierad.

Däremot kan det uppstå problem om regioner avstår från ATMP, i en situation där användning av ATMP är medicinskt och hälsoekonomiskt motiverad, därför att regionen tar hänsyn till sannolikheten att patienten med viss sannolikhet kommer att flytta. Alternativt att regioner på annat sätt försöker flytta stora engångskostnader mellan sig och detta går ut över patienterna.

Eftersom ATMP är en ny teknik under snabb utveckling råder stor osäkerhet om de långsiktiga medicinska effekterna. Kanske är det så att effekterna inte blir lika gynnsamma eller långvariga som förväntat. Det råder också osäkerhet om huruvida det kommer produkter med ännu bättre effekter i en nära framtid, som patienten kunde ha fått istället. Dessutom kan det vara så att olika bedömningar görs över tid om en och samma patient, så att två eller flera ATMP används efter varandra, eller att behandling med traditionella läkemedel återupptas efter behandling med ATMP.

Den effektiva patenttiden för läkemedel brukar uppskattas till 10-15 år.³ Efter det blir det möjligt att sälja generiska läkemedel, vilket vanligtvis innebär att priset faller kraftigt. Om priset för en engångsbehandling med ATMP ska jämföras med den förväntade livslånga kostnaden för ett traditionellt läkemedel, måste hänsyn tas till att det traditionella läkemedlet inledningsvis säljas till ett högt pris, men så småningom till ett lägre pris.

Om det är svårt att framställa generiska kopior av ATMP kommer kostnaden för framtida patienter att förbli hög även efter patenttiden. Ur ett systemperspektiv har i så fall traditionella läkemedel en fördel som inte fångas upp om kostnadsjämförelsen bara görs per patient.⁴ Se vidare kapitel 5 för en vidare diskussion.

³ Se till exempel Scott Morgon och Boller, 2017.

⁴ Scott Morton och Boller, 2017, redovisar statistik enligt priserna för *biologics* fortsätter att stiga efter patentutgång och vilken även visar att andelen för denna typ av läkemedel av de totala läkemedelskostnaderna i USA ökar snabbt.

Chapman med flera (2020) diskuterar i en OECD-rapport de utmaningar som allt dyrare cancerläkemedel resulterat i.⁵ Nya läkemedel lanseras till allt högre priser (även om läkemedelskostnaderna som andel av BNP varit relativt konstant). Nya läkemedel godkänns snabbt, ofta i "snabbspårsprocesser", eftersom behovet av behandlingar är så stort, vilket leder till stor osäkerhet om de långsiktiga hälsovinster. Dessutom utvidgas användningsområdet för de nya läkemedlen ofta successivt till indikationer där hälsovinsten är lägre och de används ofta i kombinationsbehandlingar, vilket gör att den ursprungliga hälsoekonomiska beräkningen är svår att tillämpa.

Medianpriset för nya cancerläkemedel i USA har, enligt studien, ungefär tredubblats mellan perioden 2000-2005 och perioden 2015-2019. Kostnaden för sjukhusläkemedel stiger betydligt snabbare än den för receptläkemedel i OECD-länderna.⁶ Detta gäller troligen även om det är vanligt att köparna beviljas rabatter som inte är offentliga och trots att många länder använder avtalskonstruktioner som gör att priset sänks om och när volymen ökar. Studien refererar tidigare studier, som funnit att medan generika ofta lanseras med upp till 80 procent lägre pris än det tidigare patentskyddade läkemedlet, så är priset för *biosimilars* vanligen bara 15-30 procent lägre.

Scott Morton och Boller 2017 diskuterar *biologics* i USA från liknande utgångspunkter. De visar att kostnadsandelen för *biologics*, av totala läkemedelskostnader, ökat från 15 till 25 procent mellan 2007 och 2015. Enligt författarna beror detta både på ökade kostnader för nya läkemedel och på att konkurrensen från *biosimilars* inte är effektiv. Medan traditionella läkemedel faller kraftigt i pris vid patentutgång tenderar priset på *biologics* att fortsätta att stiga. Författarna menar att detta i stor utsträckning beror på att läkemedelsbolagen lärt sig att utnyttja regelverket kring läkemedel, bland annat de regler som syftar till att stimulera utveckling av läkemedel för ovanliga sjukdomar, så kallade *orphan drugs*.

Zhai med flera (2019) redovisar att *biologics* numera står för närmare 40 procent av läkemedelskostnaderna i USA. Medan generiska läkemedel står för 90 procent av receptläkemedel i USA har få *biosimilars* trätt in på marknaden och de få som har trätt in har ofta bara lyckats ta 15-20 procent av marknaden. Förklaringarna uppges vara aggressiva juridiska domstolsprocesser för att förhindra inträde i kombination med uppgörelser som frivilligt förhindrar inträde ("*pay-to-delay*"), samt avtal med försäkringsbolag eller de organisationer som å försäkringsbolagens uppdrag förhandlar med och/eller avtalar med läkemedelsbolag om vilka läkemedel som ska erbjudas anslutna patienter. Sådana avtal kan exempelvis innebära rabatter eller bonus till dessa mellanhandsorganisationer, vilka gör att det är lönsamt för dem att rekommendera dyra läkemedel.

Grabowski med flera förutspådde en utveckling med svagt konkurrenstryck från *biosimilars* redan i ett papper från 2007. Utifrån en teoretisk analys och med hjälp av en empirisk inträdesmodell skattad med data för vanliga generika och traditionella läkemedel, förutsåg de att en försäljningspotential som kan attrahera exempelvis nio generika bara skulle kunna attrahera två *biosimilars*, om de fasta kostnaderna är 2,5 gånger så höga för de senare.

⁵ Se även Gronde med flera, 2017, som även de genomfört en litteraturstudie, i deras fall med en bredare frågeställning, vilken refereras nedan. Även de berör de problem som högt prissatta precisionsläkemedel medför.

⁶ Barlas, 2015, menar att nya cancerläkemedel ofta bara är marginellt bättre än tidigare läkemedel, medan Howard med flera, 2016, bedömer att de bidragit påtagligt till längre överlevnad, men att de ofta prissätts så högt att de inte medger någon nettohälsovinst.

Detta skulle, enligt skattningarna, leda till generikapriser som är mindre än hälften så höga som originalets men priser för *biosimilars* som är högre än 80 procent av originalets.⁷

Mestre-Ferrandiz med flera (2016) argumenterar för att skattebetalarna bäst får värde för pengarna genom att stimulera konkurrens från *biosimilars*, snarare än genom att prisreglera *biologics*. Rémuzat med flera (2017) studerar vilka faktorer som leder till hög användning av *biosimilars* i Europa, både genom litteraturstudier och genom statistisk analys. De finner exempelvis att användningen av *biosimilars* inte verkar gynnas av låga relativa priser för kopiorna.

Seoane-Vazquez med flera (2019) redovisar antalet godkända ATMP i EU och USA vid slutet av år 2018, till 13 respektive 16, samt kommenterar de höga priserna, vilka är högre i USA. Författarna bedömer att konkurrenstrycket från *biosimilars* kommer att förbli lägre än det från generika för traditionella läkemedel.

Frank och Zeckhauser (2017) diskuterar problemet att det statliga programmet Medicare i vissa delar har en utformning som gör kostnadsmotståndet för nya läkemedel svagt. De argumenterar för att detta motstånd bör stärkas och för att detta sammantaget kommer att vara samhällsekonomiskt effektivt, även med hänsyn till något svagare incitament för forskning. Detta mot bakgrund av att det finns en del som tyder på att många nya läkemedel bara medför marginella förbättringar jämfört med befintliga alternativ, eller inga förbättringar alls, trots mycket höga priser. Författarna noterar att det finns mycket litet forskning som adresserar frågan om incitamenten för forskning om och utveckling av nya läkemedel är lagom starka – även om det finns stora mängder forskning om FoU inom läkemedelssektorn.

⁷ En samtida studie av biosimilars, med fokus på regelverket för registrering och godkännande, är Kresse, 2009. En något senare litteraturstudie, Mulcahy med flera, 2014, sammanfattar de ganska optimistiska estimat av besparingspotential som var tillgängliga vid den tiden.

3 Utgångspunkter för värdering av läkemedel och institutionella förhållanden

Vår utgångspunkt är att samhället önskar maximera nyttan av läkemedel för svenska patienter per skattekrona, dock med hänsyn till att Sverige ska bidra till företagens incitament till FoU och med hänsyn till vårdens etiska principer. I det här avsnittet redogör vi för hur vi uppfattar att hälsoekonomiska bedömningar av läkemedel görs i Sverige. Den läsare som är bekant med dessa metoder kan hoppa över avsnittet eller läsa det helt översiktligt.

Det finns starka skäl för att enskilda individer inte ska drabbas av alltför höga läkemedelskostnader. I princip kan detta hanteras antingen genom försäkringslösningar eller genom offentlig finansiering. När kostnaden för läkemedel inte längre faller direkt på konsumenten riskerar dock läkemedelspriserna att bli alltför höga, i varje fall under patenttiden. Det behövs någon form av mekanism för att begränsa användning av läkemedel som det offentliga inte anser vara kostnadseffektiva, exempelvis en reglering av högsta tillåtna pris eller mekanismer för att begränsa användningen av dyra läkemedel.

En tolkning av regelverket för läkemedelsprissättning är följande. I ett första steg har samhället bestämt att patent gäller för nya läkemedel på ett sådant sätt att patentskydd föreligger i, vanligen, 10-15 år. Detta har bedömts innebära en lämplig avvägning mellan incitament till forskning och utveckling – och samhällets behov av billiga läkemedel, undvikande av dödviktsförluster och möjlighet för andra bolag att använda tidigare patentskyddade produkter i en fortsatt utveckling av nya produkter.

I ett andra steg uttrycker samhället sin maximala betalningsvilja för patentskyddade läkemedel. Om läkemedelsbolaget sätter ett pris som inte överstiger denna betalningsvilja inkluderas receptförskrivna läkemedel i förmånssystemet och läkemedel för användning på sjukhus kan upphandlas.

I ett tredje steg kan samhället vilja pressa priset på patentskyddade läkemedel under den maximala betalningsviljan för patentskyddade läkemedel. Detta kan, för sjukhusförbrukningen, ske genom upphandling i konkurrens eller genom prispförhandlingar. För receptläkemedel accepteras vanligen ett pris vid eller i närheten av maximal betalningsvilja, men ibland vidtas åtgärder för att förmå läkare att välja billigare alternativ.

Vår analys handlar huvudsakligen om det andra steget, att fastställa samhällets maximala betalningsvilja för nya läkemedel av typen ATMP. I huvudsak tar vi patentregler och incitamenten att bedriva forskning och utveckling för givna, men vi berör frågan något litet. I något större utsträckning behandlar vi det tredje steget, metoder för att pressa priset under maximal betalningsvilja.

3.1 Institutionella förhållanden

Läkemedelsbolagen är i princip fria att sätta vilket pris de vill för nya läkemedel på den svenska marknaden. När läkemedlet godkännts av Läkemedelsverket eller den europeiska godkännandemyndigheten EMA, för behandling görs dock bedömningar av huruvida behandlingen är kostnadseffektiv eller inte och olika mekanismer tillämpas för att begränsa kostnaden. Tillvägagångssätten är lite olika för receptförskrivna läkemedel respektive

läkemedel som används inom sjukhusvården, vilket i sin tur ger läkemedelsbolagen incitament att agera på delvis olika sätt.

Vi uppfattar, som nämnts, att regelverk och praxis ger läkemedelsbolagen incitament att sätta priserna för de receptförskrivna läkemedlen så nära gränsen för kostnadseffektivitet som möjligt – och ibland till och med över den gränsen. För läkemedel som används vid sjukhus uppfattar vi att förhållandena varierar mer, så att bolagen ibland har incitament att erbjuda priser som just precis är kostnadseffektiva, ibland högre än så, men i andra fall med god marginal under gränsen för kostnadseffektivitet. Särskilt för läkemedel där den totala budgetpåverkan blir stor – det vill säga för dyra läkemedel som används av många patienter – uppges regionerna ha en betydande priskänslighet, varför bolagen är benägna att erbjuda priser som ibland ligger långt under gränsen för kostnadseffektivitet.

Rekvissionsläkemedel

För läkemedel som används vid sjukhus är huvudmetoden offentlig finansiering via regionernas budget. Anskaffning sker normalt sett genom offentlig direktupphandling, i enlighet med Lagen om offentlig upphandling. Läkemedlen upphandlas av regionerna, varför priserna kan bli olika i olika regioner och varför leverantörerna möter många köpare. Flera regioner kan dock genomföra upphandlingar tillsammans. En bedömning av läkemedlets kostnadseffektivitet görs av NT-rådet, med underlag från TLV. NT-rådets bedömning är inte bindande för regionerna; regioner kan välja att inte upphandla läkemedel som NT-rådet bedömt som kostnadseffektiva.

TLV, som nationell expertmyndighet, har en rådgivande roll i förhållande till NT-rådet och regionerna och kan ge stöd inför förhandlingarna om priser (eller egentligen rabatter). TLV har ingen formell befogenhet i förhållande till priserna på sjukhusläkemedel, men myndighetens hälsoekonomiska analyser och beräkningar kan användas av regionerna vid deras beslut om vilka läkemedel som ska upphandlas och vilka priser som kan accepteras.

Läkemedelsbolagen bestämmer priser, såväl för läkemedel med patent som för läkemedel vars patent löpt ut. Vi fokuserar här på läkemedel med giltigt patent. För dessa läkemedel är det vanligt att regionerna förhandlar med läkemedelsbolagen om rabatter, snarare än om priser. I det följande resonemanget bortser vi ibland från den aspekten och talar om "pris", snarare än om "rabatt" eller "pris efter rabatt". Bolagen uppges vara benägna att erbjuda lägre effektiva priser genom rabatter, eftersom dessa inte blir publika och därmed inte påverkar prissättningen gentemot andra köpare på samma sätt.

Regionerna kan vidare på olika sätt påverka läkarna att använda kostnadseffektiva läkemedel, även i de fall dyrare alternativ har upphandlats.

Receptbelagda läkemedel

För receptbelagda läkemedel gäller andra regler.

Den metod som används i Sverige för receptförskrivna läkemedel med patentskydd är en indirekt form av prisreglering. Läkemedelsbolagen är alltså i princip fria att sätta vilket pris de vill, och läkare är fria att förskriva dem. För att få ingå i den offentliga finansieringen - läkemedelsförmånen - krävs dock att TLV bedömer att läkemedlet är kostnadseffektivt. Dessutom styr det offentliga på andra vis, exempelvis genom att utfärda rekommendationer om vilka läkemedel läkare bör förskriva.

När ett traditionellt läkemedels patent löper ut sjunker vanligen priset kraftigt genom att generiska kopior lanseras på marknaden. Vi har ovan refererat den empiriska litteraturen om läkemedelspriser vid patentutgång. Enligt denna faller priserna inte alls lika mycket för biologiska läkemedel (*biologics*), till vilka ju ATMP hör, vid patentutgång.

För receptförskrivna läkemedel som inte längre åtnjuter patentskydd eftersträvas kostnadsbaserade priser. Den metod som används för detta är ett månadsvist auktionsliknande förfarande som pekar ut det för tillfället billigaste generiska alternativet, i kombination med att den offentliga finansieringen av läkemedel bara täcker en summa motsvarande det billigaste alternativet. Överskjutande del av priset betalas normalt av patienten. Förenklat blir effekten av dessa regler, som nämnts ovan, att priserna för receptläkemedel förblir nära gränsen för kostnadseffektivitet under hela patenttiden, för att därefter falla kraftigt.

Eftersom läkemedel för användning på sjukhus anskaffas genom offentlig upphandling och eftersom regionerna uppfattas vara mer priskänsliga än staten finns bättre förutsättningar att få till stånd prispress redan under patenttiden. Detta gäller i synnerhet om det finns flera läkemedel som kan användas för samma indikation, men enligt ovan också för läkemedel där den totala budgetpåverkan riskerar att bli stor. Samtidigt gäller att patentutgång inte på samma sätt utlöser hård prispress för ATMP och andra *biologics*.⁸

Sammantaget innebär den värdebaserade prissättningen och generikasystemet att samhällets kostnader för läkemedel som säljs mot recept begränsas, men också att Sverige bidrar till att företagen får ekonomiska incitament att bedriva forskning i syfte att ta fram nya läkemedel.

3.2 Samhällsekonomisk kostnads-intäktsanalys

I en samhällsekonomisk kostnads-intäktsanalys, som en hälsoekonomisk utvärdering är ett exempel på, ställs på ett systematiskt sätt samhällets intäkter eller nytta av ett visst handlingsalternativ mot samhällets kostnader för detsamma. I en sådan analys ska hänsyn tas till samtliga kostnader och intäkter, inte bara rent monetära. Exempelvis ska negativa effekter på miljö och hälsa räknas som en kostnad, även om de inte dyker upp som en kostnad som belastar en budget, och på samma sätt ska en icke-monetär nytta räknas som en intäkt. I hälsoekonomiska kalkyler är bättre hälsa i form av ökad överlevnad och bättre livskvalitet de viktigaste intäkterna, medan viktiga kostnadsposter är vård- och läkemedelskostnader.

Eftersom kostnader och intäkter ofta är utspridda i tiden behövs ett sätt att göra dem jämförbara. Den metod som används är diskontering, vilket behandlas utförligt i en separat underlagsrapport skriven av Johan Stennek.

En kostnads-intäktsanalys kan användas för att avgöra om det är samhällsekonomiskt motiverat, eller inte, att följa ett visst handlingsalternativ. Det beslutskriterium som normalt används i denna typ av situation är nuvärde. Handlingsalternativet väljs, vilket exempelvis kan innebära att en investering genomförs, om nuvärdet är positivt.

⁸ En observation är att prismekanismerna för sjukhusförbrukningen mer liknar hur det fungerar i USA, med många olika aktörer och där dessa aktörer var och en för sig kan välja att utesluta särskilt dyra läkemedel.

Kostnads-intäktsanalys kan också användas för att välja mellan flera olika handlingsalternativ med positivt nuvärde. I detta fall är det mer komplicerat att säga vilket beslutskriterium som ska användas, eftersom det beror på situationen. Är de olika handlingsalternativen ömsesidigt uteslutande eller kan alla eller flera genomföras samtidigt? Är handlingsalternativen - till exempel investeringarna - fritt skalbara eller har de en viss given storlek?

3.3 Värdering av liv och hälsa

Eftersom samhället inte har obegränsade resurser för hälsovård, är vi på något sätt tvungna att välja vilka insatser som ska utföras. Det innebär konkret att hälsoeffekter på något sätt måste vägas mot kostnader. Detta görs med hjälp av *kostnads-effektivitetsanalys*, där de insatser som kan göras till lägst kostnad per hälsovärde väljs, där samhället på något sätt har valt hur hälsoeffekter ska kvantifieras.⁹

I bedömningen av läkemedlets kostnadseffektivitet använder TLV så kallad *värdebaserad prissättning*. Den innebär att samhällets betalningsvilja för ett läkemedel ska baseras på det värde som läkemedlet skapar i förhållande till det värde som skapas av bästa befintliga behandling. Det värde som skapas är bättre hälsa och ökad livslängd, vilket uttrycks i QALY, (Quality Adjusted Life Years), samt genom samhällets betalningsvilja för en QALY. Den värdebaserade prissättningen beskrivs mer utförligt nedan.

I princip sätts ett ekonomiskt värde på ett levnadsår i full hälsa, det vill säga en QALY. Ett år med påtagligt reducerad hälsa, till exempel svår smärta, åsätts ett reducerat värde. Reduktionen ska vara proportionell mot hur allvarligt hälsan är reducerad.

Det kan tyckas svårt att sätta ett värde på ett liv eller ett levnadsår. Etiska principer kan tyckas kräva att ett liv ska värderas oändligt högt. Principen kan dock försvaras med i synnerhet två argument. För det första är samhällets resurser begränsade. Vi har inte oändliga resurser att använda för att rädda liv. Samhället måste prioritera de insatser som är mest effektiva för att rädda liv eller att förbättra levnadsförhållanden. För det andra förefaller det som att människor, så länge riskerna är måttliga, agerar som om det värde de sätter på sitt eget liv är begränsat. Exempelvis väljer flertalet människor att åka bil och flyg, trots att det medför en viss dödsrisk och väljer ibland bil framför kollektiva färdmedel, trots att dödsrisken i det förra fallet är högre.

Vid värderingen intar samhället rollen som den representativa medborgaren/patienten och ställer sig bildligt talat bakom en slöja av okunnighet. Den representativa medborgaren kommer med viss sannolikhet att drabbas av olika sjukdomar, olyckor och för tidig död. Samhället har kunskap om riskerna för dessa händelser, om sannolikheten att olika mediciner leder till hälsovinster, samt om samhällets ekonomiska resurser. Hänsyn tas till hela livsrytmen.

Vid värderingen läggs dock vissa moraliskt grundade principer på den värdering som i princip skulle kunna tas fram genom att ett stort antal människor tillfrågas. Exempelvis tillämpas samma värde per levnadsår i full hälsa, oavsett ålder och inkomst.

⁹ Vi diskuterar nedan de problem som kan uppstå när vissa handlingsalternativ är ömsesidigt uteslutande. Om sådana finns är rangordning av alternativ enligt kostnadseffektivitet inte nödvändigtvis rätt beslutskriterium.

Att ett levnadsår med begränsad hälsa värderas lägre än ett levnadsår med fullgod hälsa kan i det sammanhanget framstå som oetiskt. Dock är detta närmast nödvändigt för att på ett logiskt konsekvent sätt kunna värdera läkemedel och sjukvård. I motsatt fall, om ingen skillnad görs mellan ett år med god respektive dålig hälsa, skulle konsekvensen bli att ett läkemedel som återställer full hälsa inte är något värt.

Ett etiskt dilemma uppstår för mycket dyra läkemedel med en direkt livräddande effekt för vissa identifierade patienter. Samhället kan inta en sparsam och strikt hållning när det gäller läkemedel som bara reducerar en redan låg eller måttlig dödsrisk något. Och kanske kan samma hållning intas när det gäller läkemedel som bara med en viss låg sannolikhet hjälper en allvarligt sjuk eller döende individ. Men det är svårt eller omöjligt att av ekonomiska skäl avstå från att hjälpa en döende individ som med säkerhet eller näst intill säkerhet kan räddas till en väldigt hög kostnad, enligt *The Rule of Rescue*. Däremot kan samhället i förväg välja att dimensionera sin kapacitet att rädda liv på ett sådant sätt att den förväntade kostnaden per vunna QALY blir rimlig. Exempelvis väljer samhället att ha ett begränsat antal respiratorer, akutmottagningar och läkare. I ett läge när all personal och alla faciliteter redan används för lika akuta ändamål – till exempel i ett katastrofläge – hjälper det inte att åberopa *Rule of Rescue*.

Läkemedelsbolag kan exploatera det moraliska dilemman genom att kräva väldigt höga priser för läkemedel som används på ett sådant sätt att *Rule of Rescue* kan åberopas. Detta gäller för läkemedel som används för vissa patientgrupper där det inte finns goda alternativ och där det inom kort tid och med stor säkerhet uppstår allvarliga hälsoeffekter om läkemedlet inte används. Om sjukvården då inte betalar det höga priset utsätts de för påtryckningar från patienter och anhöriga.

3.4 Sjukvårdens budget

Sjukvårdens budgetrestriktion kan påverka bedömningen av ATMP på flera sätt. Ett större budgetutrymme gör det möjligt att acceptera läkemedel med högre kostnad per QALY, vilket kan förväntas bli konsekvensen av att vårt samhälle successivt blir rikare. Samtidigt finns prognoser som pekar mot att det i framtiden kommer att bli svårare och svårare att finansiera stigande vårdkostnader, som en konsekvens av nya dyra läkemedel, ny teknik, en åldrande befolkning och på grund av svårigheterna att rationalisera tjänsteintensiv produktion (Baumol-effekten).¹⁰

En budgetrestriktion kan vara hård eller mjuk. Med en hård budgetrestriktion kommer en tillkommande utgift att tränga undan en annan lika stor utgift; med en mjuk restriktion sker undanträngningen bara delvis eller inte alls. Om budgetrestriktionen för sjukvårdens läkemedel är hård och det tillkommer stora kostnader för exempelvis ATMP, kommer en betydande undanträngning att ske. I princip trängs de minst kostnadseffektiva läkemedlen undan först, därefter successivt allt mer kostnadseffektiva. Gränsvärdet för den kostnad per QALY som accepteras justeras ner i takt med denna undanträngning och en lägre Igräns måste följaktligen tillämpas på tillkommande ATMP.

Ett sätt att tänka kring budgetrestriktionen är ställa sig frågan om kostnadsramen för läkemedel är förutbestämd och bara de (mest kostnadseffektiva) läkemedel som ryms inom den givna kostnadsramen får användas – eller om patienternas rätt till läkemedel har

¹⁰ Lorenzi et al, 2019,

definierats och kostnaden bestäms av hur många patienter som behöver de läkemedel de har getts rätt till. Med ekonomjargong är budgeten i det första fallet exogen eller diskretionärt bestämt; i det senare fallet är budgeten endogen eller förmånsbestämd.

En kostnad kan alltså vara tvingande (*mandatory spending*, enligt amerikansk terminologi) eller diskretionär (*discretionary spending*). Låt säga att TLV tar fram ett beslutsunderlag där det slutliga priset som erbjuds tillsammans med hälsovinsten räknas om till en kostnad per QALY. Om kostnaden per QALY är tillräckligt låg för ett receptläkemedel tas det in i förmånssystemet; om den är tillräckligt låg köper regionen in läkemedlet till sjukhusen. Tröskelvärdet för kostnad per QALY är stabil från år till år och antalet nya läkemedel som köps in respektive subventioneras bestäms av marknaden. Det uppstår en tvingande kostnad, det vill säga regionen respektive staten har inte särskilt stor kontroll över läkemedelskostnaden.

Om kostnaden för sjukhusförbrukningen mer är att betrakta som diskretionär, kan i princip läkemedel som sjukhuset (eller dess läkare) önskar använda under en viss period, exempelvis ett år, sättas upp på en lista, där läkemedlen rangordnas så att det med lägst ICER står först, därefter det med näst högst ICER osv. Regionen eller sjukhuset använder läkemedel så långt ner på listan som den (diskretionärt) tilldelade budgeten räcker. Patienter som hamnar ”under strecket” får vänta till nästa period (år) och hoppas att strecket då dras längre ner.¹¹

Vi uppfattar att beslutsprocessen inom läkemedelsförmånen i huvudsak svarar mot en endogen läkemedelsbudget. Vid introduktionen av ett nytt läkemedel gör TLV en bedömning av dess kostnadseffektivitet. Tröskelvärdena ändras endast långsamt och i varje fall på kort sikt tas besluten utan hänsyn till prognosticerade kostnader i förhållande till budget.

För sjukhusförbrukningen finns en tydligare koppling mellan budgetutfall och betalningsvilja. Den mängd vård som utförs, i bemärkelsen läkarbesök, operationer och andra behandlingar, kan vara kvantitativt begränsad av sjukvårdssystemets kapacitet och budgetutrymme. Vården ransoneras genom köer i kombination med vårdprioriteringar enligt angelägenhetsgrad. I någon mån kan användningen av ATMP och andra dyra läkemedel begränsas av den här typen av mekanismer. Vi uppfattar att när väl ett ATMP godkänts för användning i en region kommer användningen inte att vara begränsad av en regionens vårbudget eller behandlingskapacitet utan i princip användas för alla patienter som har medicinsk nytta av läkemedlet.¹² Däremot uttrycks regionernas kostnadsmotstånd genom att de tvekar att upphandla särskilt dyra läkemedel. Detta gör i sin tur läkemedelsbolagen mer mottagliga för argument om att kostnaden inte ryms inom budgeten – och därmed mer förhandlingsbenägna.

De två budgeteringsmodellerna har för- och nackdelar. Att regionen delvis tappar kontrollen över kostnaderna är förstas är en nackdel med *mandatory spending*, särskilt om annan vård som exempelvis operationer trängs undan. En fördel är dock att patienter vad gäller tillgången till läkemedel behandlas lika från ett år till ett annat och från en region till en annan, om alla regioner använder samma gränsvärde för ICER. En annan fördel är att TLV:s

¹¹ Denna beslutsalgoritm är dock inte alltid optimal, vilket diskuteras vidare nedan.

¹² I praktiken kommer förstas ATMP att introduceras gradvis, men av andra skäl än rent ekonomiska eller kapacitetsmässiga.

praxis kring gränsvärden för kostnad per QALY kan användas mer eller mindre "rakt av" – exempelvis praxis som säger att en högre kostnad kan accepteras för allvarliga sjukdomar.

Sammanfattningsvis påverkas prissättning och hälsoekonomiska bedömningar av ATMP på flera sätt av läkemedelsbudgeten och dess utveckling. För det första förväntas samhället bli allt rikare, vilket talar för att det är bra att skjuta utgifter på framtiden. För det andra ändras tröskelvärden för ICER över tiden, exempelvis beroende på om det finns stora utgiftspucklar för exempelvis ATMP, men även av andra skäl. Om samhället ett visst år står inför en ovanligt stor kostnad för ATMP och om sjukvårdens budgetrestriktion är hård, bör gränsvärdet för ICER sänkas det året, såväl för ATMP som för andra läkemedel och åtgärder inom sjukvården. För det tredje och relaterat till föregående punkt leder en hård budgetrestriktion till ett hårdare kostnadsmotstånd från regionerna och därmed en större förhandlingsbenägenhet från läkemedelsbolagens sida.

3.5 ICER, Net Benefit och marginalnyttan av inkomst

När det finns många insatser som kan utföras av vården, vilka insatser ska väljas? Det finns olika vårdbehov och vi vill uppnå så stor ökning som möjligt av livskvalitet med en given budget. Som huvudregel prioriteras de insatser som innebär störst ökning av hälsorelaterad livskvalitet per krona högst. Insatserna rangordnas i fallande ordning från högst värde per krona till lägst och de som ryms inom budgetutrymmet utförs. Om de mest angelägna åtgärderna som inte ryms inom budget ändå framstår som mycket angelägna, finns anledning att utöka budgeten.

I praktiken hanteras sjukvårdens kostnader inte på ett så rationellt och mekaniskt sätt. Vad som ryms inom budget påverkas också av politiska överväganden på olika nivåer, sannolikt också av interpolitik, liksom av traditioner och etiska och professionella bedömningar. För vårt ändamål är dock detta en användbar modell eller tankeram.

Analysen är i princip densamma oavsett om sjukvårdsbudgeten är optimalt bestämd i en dynamisk process eller fix, det vill säga oberoende av om budgeten är hård eller mjuk med den terminologi vi använde i föregående avsnitt. Den princip för prioritering som vi redogjort för gäller så länge insatserna är oberoende.

Om däremot de olika insatserna är alternativ, det vill säga ömsesidigt uteslutande, kan det finnas anledning att välja på ett annat sätt. En dyr behandling kan ge en större total netto nytta men vara mindre kostnadseffektiv än en mindre dyr behandling. Om detta gäller bör ofta den behandling som ger högst total netto nytta prioriteras.

För ATMP gäller generellt att såväl nytta som kostnad är höga, vilket gör att nytta/kostnadskvoten möjligen inte är särskilt hög, men det är ändå möjligt att totalt nuvärde är högt. Denna möjlighet diskuteras vidare i avsnitt 3.8

Vi återgår till antagandena att budgeten är begränsad och att de behandlingar som kan utföras är oberoende av varandra. När en ny behandling ska bedömas räcker det med att jämföra med den marginella behandling (den minst kostnadseffektiva) som ryms inom budgeten, eftersom det är den som riskerar att trängas undan.

Om den tillkommande behandlingen har en låg kostnad behöver inte tröskelvärdena justeras. Om den tillkommande behandlingen har en hög kostnad kommer tröskelvärdet för QALY att behöva justeras ned och gränsvärdet för ICER behöver sänkas. Vissa behandlingar

som tidigare utfördes trängs undan – eftersom vi antar att budgeten är begränsad. Se appendix A för en mer teknisk beskrivning.

3.6 Värdebaserad prissättning

Den metod som används i Sverige för att bedöma värdet av läkemedel, värdebaserad prissättning, är en så kallad kostnads-effektivitetsanalys.¹³ Antag att en beslutsfattare ska välja vilka behandlingar som bör utföras av ett antal (oberoende) alternativ med olika QALY och kostnader för dessa. Ett optimalt sätt att välja är, enligt ovan, att värdera kostnaden per QALY för alla behandlingar och att utföra de med lägst kostnad inom ramen för den budget som finns. Eftersom det leder till lägre kostnad för varje nyttoenhet kan mer nytta uppnås inom ramen för hälsobudgeten. Om hälsobudgeten inte är helt given är det optimalt att välja alla behandlingar som har lägre kostnader per nytta än andra användningar i samhället (t.ex. privat konsumtion eller investeringar).¹⁴

En annan slutsats av denna analys är att om ett nytt läkemedel prissätts på så sätt att dess ICER ligger bara litet under tröskelvärdet, då är också ökningen i samhällsnytta av att använda behandlingen bara litet större än den behandling som trängs undan på marginalen.

För ett läkemedel som söker tillträde till det svenska förmånssystemet görs en utredning av de hälsovinster, mätta i QALY, som läkemedlet medför. Detta görs antingen i förhållande till ingen behandling, om det inte finns någon sådan, eller i förhållande till bästa alternativa behandling, ”jämförelsealternativet”, om sådan finns. Om det finns ett jämförelsealternativ beräknas hur många ytterligare QALY som skulle vinnas, när det nya läkemedlet istället används. Denna inkrementella nytta ställs mot den inkrementella kostnaden, det vill säga merkostnaden för det nya läkemedlet jämfört med det gamla. Om nettot (*net benefit*) är positivt bör läkemedlet beviljas tillträde till förmånssystemet. Resultatet beror förstås på vilket värde som sätts på en QALY; ett högre värde gör att högre läkemedelskostnader kan accepteras och omvänt.¹⁵

Att ett nytt läkemedel ställs mot ett jämförelsealternativ är förenligt med de allmänna principerna för samhällsekonomisk kostnads-effektivitetsanalys. Förutsatt att det gamla läkemedlet är samhällsekonomiskt lönsamt, jämfört med alternativet ingen behandling, kommer en inkrementell analys inte att leda till att ett nytt läkemedel godkänns, som inte skulle bli godkänt om referenspunkten istället var ingen behandling. Omvänt gäller dock att nya läkemedel, som skulle vara samhällsekonomiskt lönsamma jämfört med ingen behandling, likväl kan bli underkända när de ställs mot jämförelsealternativet.

I Figur 1 är alla kombinationer av vunna QALYs och kostnader som befinner sig under och till höger om den tjocka streckade linjen samhällsekonomiskt lönsamma, jämfört med ingen

¹³ För en teoretisk presentation av värderingsmetoden, se Drummond med flera (2015). För en mer ingående teoretisk analys, se Garber och Phelps (1997).

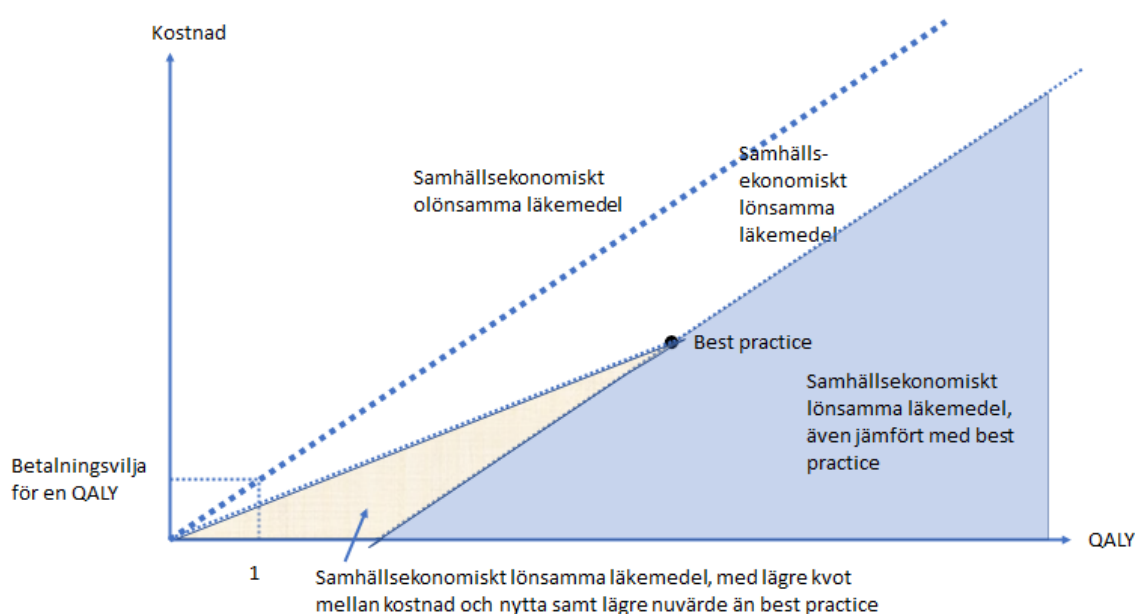
¹⁴ För att bestämma sjukvårdsbudgetens storlek i förhållande till annan användning krävs att man på något sätt jämför nyttan av hälsoinsatser med andra alternativ.

¹⁵ Det bör påpekas att TLV inte fastställer priser. TLV godkänner eller godkänner inte att ett läkemedel ingår i läkemedelsförmånen, utifrån det pris som läkemedelsföretaget föreslår och utifrån de hälsovinster företaget kan göra sannolikt uppstår vid användning av det nya läkemedlet. TLV har gjort det känt för marknaden enligt vilka principer hälsovinster värderas, vilka krav som ställs på dokumentation av de medicinska effekterna, samt en ungefärlig värdering av en QALY.

behandling. Jämfört med jämförelsealternativet är kombinationer av vunna QALYs och kostnader som befinner sig inom den blåtonade ytan samhällsekonomiskt lönsamma.¹⁶

De kombinationer av vunna QALYs och kostnader som ligger i den gula triangeln har en lägre kvot mellan kostnader och vunna QALY än jämförelsealternativet (*best practice*), men ger samtidigt ett lägre nuvärde. Vid valet mellan en sådan kombination och jämförelsealternativet bör det senare väljas. Detta är alltså ett exempel på att kvoten mellan nytta och kostnad (ICER) inte alltid är rätt beslutskriterium, vilket alltså diskuteras vidare i avsnitt 3.8.¹⁷

Figur 1. Värdebaserad prissättning, samhällsekonomisk lönsamhet och kostnads/nyttokvot



3.7 Prissättning och konkurrens

När patentet löper ut för det läkemedel som är jämförelsealternativ kan man förvänta sig att priset faller påtagligt.¹⁸ Eftersom den värdebaserade prissättningen bygger på inkrementell nytta i förhållande till inkrementell kostnad förändrar detta bedömningen även av läkemedel med fortsatt patentskydd. För nya läkemedel, som prövas efter det att priset sänkts för jämförelsealternativet, kommer det lägre priset för alternativet att slå igenom med automatik, men i princip gäller detsamma för läkemedel med patent vars pris redan godkänts. För att det också ska slå igenom i praktiken krävs dock att TLV reviderar redan genomförda bedömningar. Principen illustreras i Figur 2. Om patentutgång för jämförelsealternativet (*best practice*) är nära förestående bör detta beaktas i analysen.

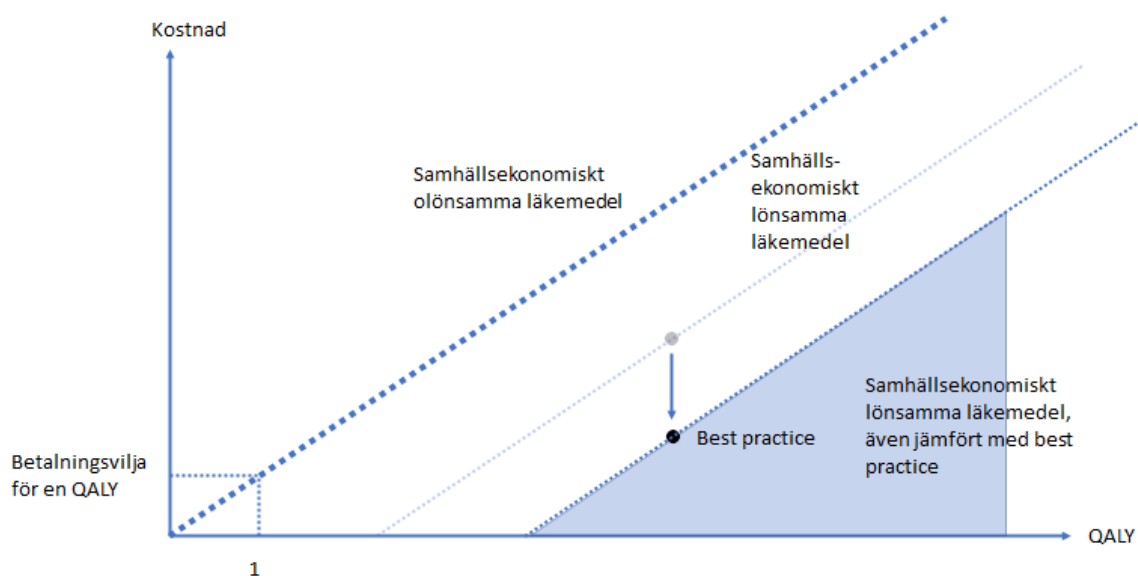
¹⁶ Vi är medvetna om att TLV vanligen illustrerar detta i ett kostnads-effekt-diagram, där jämförelsealternativet, i diagrammet kallat *best practice*, är placerat i origo.

¹⁷ Se Drummond med flera, 2015, för en utförlig diskussion av hälsoekonomisk analys och värdebaserad prissättning.

¹⁸ Som vi diskuterar på annat ställe gäller detta inte nödvändigtvis för ATMP.

Om en framtida patentutgång för jämförelsealternativet inte beaktas kan det leda till att ett nytt läkemedel vid den gråstreckade linjen godkänns, trots att det inte är kostnadseffektivt när priset på jämförelsealternativet sjunker.

Figur 2. Värdebaserad prissättning och priskonkurrens vid patentutgång



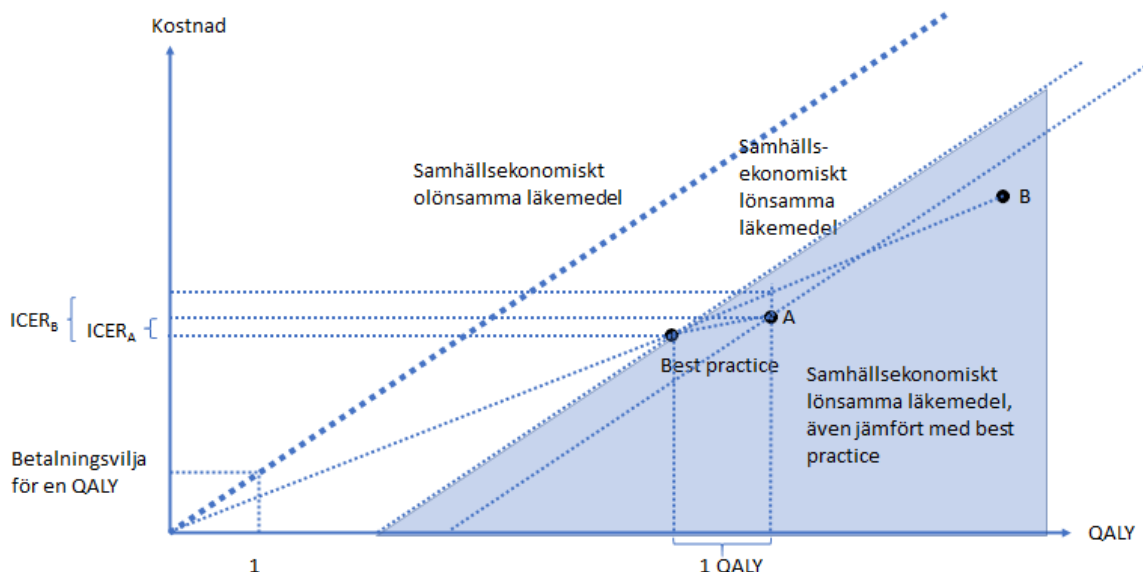
3.8 Val mellan alternativa behandlingar

Om två eller flera behandlingar är helt oberoende av varandra, till exempel för att de används för olika patienter, bör det alternativ som är mest kostnadseffektivt, det vill säga har lägst ICER, väljas. Detta har vi beskrivit ovan. Vid hälsoekonomiska analyser finns dock ofta ett beroende mellan de olika alternativen. Två läkemedel för samma sjukdom och samma patient kan kanske inte kombineras och även om de kan det, så kan nyttan antagligen inte bara adderas. I det följande antas att olika behandlingar är ömsesidigt uteslutande.

Figur 3 illustrerar att lägst ICER *inte* är rätt beslutskriterium vid valet mellan två ömsesidigt uteslutande behandlingar. Med utgångspunkt från ett jämförelsealternativ (*best practice* i figuren) uppskattas inkrementell hälsoekonomisk vinst eller nytta och inkrementell kostnad för det nya läkemedlet. I figuren finns två nya läkemedel, A respektive B. Vid användning av A uppskattas den inkrementella hälsovinsten till 1 QALY och den inkrementella kostnaden motsvarar sträckan ICER-A, som också är ICER för läkemedel A.

För läkemedel B är den inkrementella hälsovinsten större, liksom den inkrementella kostnaden. Dessa är inte markerade i figuren, men motsvarar det horisontella respektive det vertikala avståndet mellan punkten *best practice* och punkten B. Per QALY är kostnaden ICER-B, det vill säga en högre ICER.

Figur 3. ICER – inte alltid rätt kriterium för att prioritera mellan ömsesidigt uteslutande läkemedel



En rangordning enligt ICER (kostnadseffektivitet) skulle visa att läkemedel A är att föredra framför läkemedel B. Detta är dock fel, eftersom den hälsoekonomiska analysen bör eftersträva att så hög nettonytta som möjligt uppnås. Att B ger högre nettonytta framgår grafiskt genom att punkten B ligger nedanför och till höger om en linje genom punkten A, som är parallell med den linje som avgränsar det skuggade fältet uppåt och åt vänster.¹⁹

Konkret kan "lägsta ICER" vara fel beslutskriterium i en valsituation mellan ett relativt billigt läkemedel som ger hög nytta per krona och ett dyrare läkemedel som ger lägre nytta per krona, men som totalt sett ger mer nytta och som ändå är kostnadseffektivt, jämfört med många andra läkemedel. Se Drummond (2005, sid 98-106) för en mer ingående diskussion om beslutskriterier och ICER när det finns flera alternativa behandlingar.

När det gäller specifikt ATMP är de ofta just dyra läkemedel och kan ha relativt hög ICER. Om ett ATMP ställs mot en befintlig behandling ("jämförelsealternativet") och om inkrementell nytta jämförs med inkrementell kostnad medför det ingen komplikation att de två behandlingarna är ömsesidigt uteslutande. ATMP kommer att accepteras om det är kostnadseffektivt jämfört med det befintliga alternativet och ATMP kommer då att medföra en högre nettonytta.

Problemet uppstår om ICER beräknas för ett stort antal läkemedel, om dessa rangordnas efter ICER och om de läkemedel köps in som ryms inom en given budget. Med en sådan beslutsprocess kan ett läkemedel motsvarande A i figur väljas före ett läkemedel som B, fast det senare egentligen skulle ha medfört en större nettonytta för samhället.

Om samhällets (regionernas) budget för vårdutgifter eller läkemedelskostnader är "hård" kan tillkomsten av möjligheten att behandla många patienter med dyra ATMP leda till att tidigare hälsoekonomiskt lönsamma behandlingar trängs undan och att tröskelvärdet för

¹⁹ Linje med en sådan lutning kan ses som samhällsekonomiska indifferenskurvor.

ICER därför sänks. (Se diskussionen i avsnitt 3.5.) Grafiskt innebär detta att linjen som går genom *best practice*, som avgränsar behandlingar med positiv nettonyttan från de med negativ nettonyttan, kommer att vridas medurs runt *best practice* (eller jämförelsealternativet). Om läkemedelsbudgeten blir tillräckligt knapp och tröskelvärdet för ICER därför faller tillräckligt mycket, kommer så småningom behandling A att föredras framför behandling B även enligt kriteriet nettonyttan.

En annan situation som kan innebära att "lägsta ICER" leder fel är när ett läkemedel som är något mindre hälsoekonomiskt lönsamt trängs undan av ett mer hälsoekonomiskt lönsamt, men där det första läkemedlet är oersättligt och förebygger en irreversibel skada, medan det andra, i princip mer kostnadseffektiva läkemedlet, kan sättas in vid en senare tidpunkt och göra god nytta.

3.9 Patientens egenskaper (indikationsbaserad värdering)

Hittills har vi bortsett från att värdet av ett läkemedel även beror på vilken och hur svår sjukdom patienten lider av. Samma läkemedel kan tänkas bota en svårt sjuk människa och en lindrigt sjuk människa och det kan tänkas förlänga livet mer på en ung människa än på en äldre människa. Inom en diagnos görs beräkningen av hälsovinster utifrån den förväntade fördelningen av patienternas tillstånd och hälsovinster, vilket avgör hur högt pris som kan accepteras. Förväntad återstående livstid för populationen påverkar värderingen. Däremot påverkar ålder och hälsostatus, givet samma indikation, inte vilka patienter som har rätt att få läkemedlet finansierat av det offentliga.

När det gäller olika indikationer förhåller det sig annorlunda. Prissättningen av ett läkemedel godkänns, det vill säga tas in i förmånssystemet, för en viss indikation (eller för flera indikationer), där hälsovinster i förväntan är stora, men inte för en eller flera andra indikationer, där hälsovinster är mindre. På motsvarande sätt kan regionerna köpa in ett läkemedel att användas inom sjukhusen för vissa indikationer, men på olika sätt försöka begränsa eller förhindra dess användning för andra indikationer.

Ett alternativ är dock att läkemedelsbolaget får ta ett högre pris för den ena indikationen än för den andra. Detta brukar kallas indikationsbaserad prissättning. Fördelen med indikationsbaserad prissättning är att det kan leda till ett bättre utnyttjande av befintliga läkemedel (mindre dödviktstförluster). Om läkemedelsbolaget inte får ta olika priser sätter det ett högt pris, vilket leder till att läkemedlet enbart används för den ena indikationen, där det ger störst medicinskt värde. Om läkemedelsbolaget får ta olika priser kommer läkemedlet att användas för båda indikationerna.

Nackdelen med indikationsbaserad prissättning är att det kan leda till en överföring av överskott från det offentliga till företagen. Om bolaget inte får använda olika priser för samma läkemedel är det ibland mer lönsamt att sätta ett lågt pris för att sälja till båda typerna av patienter. Om olika priser tillåts kommer priset för den ena patientgruppen att höjas, vilket alltså bara överför ett överskott till företaget.

Ytterligare en fördel kan dock följa av denna nackdel. Det kan vara så att den högre lönsamheten som följer av indikationsbaserad prissättning kan vara förutsättningen för att det alls ska vara värt att utveckla läkemedlet. Det krävs en hög lönsamhet för att investeringar i FoU ska vara motiverade och för vissa läkemedel kan en sådan lönsamhet nås bara om läkemedlet får säljas för olika priser till olika patientgrupper.

Kombinationsbehandlingar, exempelvis vid cancerbehandling, är ett motiv för indikationsbaserad prissättning, se OECD (2020). Ett läkemedel A kan användas för vissa indikationer, ett annat, B, för andra, medan kombinationen är effektiv för ytterligare andra indikationer. I ett enkelt exempel är marginalkostnaden för såväl A som B lika med 0 och värdet av behandlingen är 1 för alla indikationer. Läkemedelsbolaget får ta priset 1 för de respektive indikationerna där endast ett läkemedel används. Om samma priser används är kombinationsbehandlingen inte kostnadseffektiv, vilket är ett samhällsekonomiskt slöseri.

Om läkemedlen har samma tillverkare kan ett kombinationsläkemedel tas fram, men om det är två olika tillverkare är det svårt för dem att enas om priser – och dessutom möjligen i strid med konkurrensreglerna.

4 Beräkning av kostnad per QALY för ATMP

Vid värdering av ATMP måste en jämförelse göras med alternativ behandling över lång tid. Om inte hänsyn tas till förväntad utveckling riskerar ATMP att övervärderas av framför allt tre skäl:

- *Hälsoeffekten av ATMP kan vara lägre än förväntat.*
- *Utfallet om inte ATMP väljs kan, på grund av framtida läkemedelslanseringar, vara bättre än nuvarande jämförelsealternativ.*
- *Kostnaden för jämförelsealternativet sjunker med tiden.*

Med traditionella läkemedel som ges som kontinuerlig behandling är vårdgivaren fri att byta behandling om omständigheterna förändras. Oåterkalleligheten av ATMP behandlingar kombinerat med engångsbetalning innebär att utfall långt fram i tiden måste värderas nu, för att inte övervärdera ATMP.

I och med att ATMP är en botande behandling, måste valet att använda ett ATMP baseras på ett förväntat utfall över mycket lång tid. Värderingen av ATMP måste ta hänsyn till de osäkerheter som finns, till de relevanta valmöjligheter som finns nu men också till relevanta valmöjligheter som kan uppstå i framtiden. Det innebär att man i princip måste göra långsiktiga prognoser av vad som händer vid respektive val.

Vi använder begreppet *kontrafaktiskt scenario* (eller kontrafaktiskt alternativ) för att betona att valmöjligheterna och därmed jämförelsealternativen är mer komplexa än när traditionella läkemedel ska värderas. Det spelar också roll om den sjukdom som ett ATMP behandlar är livshotande eller om den obehandlad kan orsaka irreversibla skador liksom om det i så fall finns andra läkemedel som kan användas. Vi använder begreppet *oersättlig* för ATMP som kan behandla livshotande sjukdomar för vilka det tidigare saknats behandling eller som kan förhindra att permanenta skador uppstår till följd av en sjukdom och där det inte finns traditionella läkemedel som kan fylla samma funktion.

Värderingen av läkemedel syftar till att analysera om kostnaden per QALY understiger det samhället maximalt är villigt att betala. För traditionella läkemedel sjunker kostnaden per QALY över tiden med introduktion av terapeutisk och generisk konkurrens. Vi väntar till nästa kapitel med att diskutera hur denna förväntade prisutveckling påverkar värderingen av ATMP.

Värderingen av ett nytt läkemedel är som vi diskuterade i föregående kapitel, baserad på fyra komponenter: kostnad och hälsoutfall, för den nya och för den befintliga behandlingen. Tre av dessa är osäkra, men kostnaden för den nya behandlingen bestäms vid den upphandling (eller motsvarande) som den hälsoekonomiska analysen ger underlag för.

Vi måste alltså prognosticera:

1. Förväntad QALY för ATMP-behandling
2. Kostnad för annan behandling
3. Förväntad QALY för annan behandling

I den långsiktiga bedömningen av ATMP är det viktigt att utreda vad som skulle hända om ATMP inte användes, det som ibland kallas det *kontrafaktiska* alternativet. När det gäller

kostnader och hälsoutfall utan ATMP-behandling är det komplicerat, men för mycket allvarliga sjukdomar är värderingen paradoxalt nog lättare. Om patienter relativt snart dör utan behandling med ATMP, innebär det också att prognostiden är kort och osäkerheten liten för punkterna 2 och 3 ovan.

För sjukdomar där patienter som inte får ATMP ändå överlever länge, med eller utan annan behandling, är utfallet osäkrare eftersom prognostiden är längre. För att göra en väntevärdesriktig jämförelse måste man dock göra en prognos för det kontrafaktiska alternativet. Om inte ATMP hade funnits, vad hade vi förväntat oss för hälsoeffekt – inklusive kommande innovationer?²⁰ Kommande innovationer i annan behandling leder till att hälsovinsten eller kostnadsbesparingen med att välja ATMP för de patienter som behandlas blir mindre.

I detta kapitel analyserar vi hur värdering av ATMP är annorlunda än för traditionella läkemedel. Utgångspunkten är att leverantören av ATMP ersätts med en engångsbetalning vid behandlingstillfället. Om man kan välja en betalningsmodell där betalningen baseras på utfall, kan osäkerheter i dessa utfall hanteras i viss mån. I kapitel 7 diskuterar vi sådana möjligheter. Även om utfallsbetingad prissättning (betalning) är möjlig, är det viktigt att studera hur prissättning under osäkerhet skulle formuleras. Att det råder osäkerhet om de avtalsmässiga och praktiska förutsättningarna för utfallsbetingad prissättning gör detta än viktigare.

4.1 Allmänna utgångspunkter

4.1.1 Väntevärdesriktig värdering

Osäkerheten med ATMP är viktig att beakta, framför allt i och med att den i stor utsträckning inte är symmetriskt fördelad. Om det är mycket mer sannolikt att den inkrementella kostnaden per QALY överstiger det prognosticerade utfallet än att den understiger den, utgör inte det prognosticerade värdet det förväntade utfallet. Det förväntade hälsoekonomiska värdet av ATMP kommer då systematiskt att överskattas. I och med att värderingen görs över en lång tidshorisont kan även små sannolikheter få stora effekter på värderingen.

De tre typerna av osäkerhet (eller förväntningar) vi behöver beakta är alltså:

De framtida hälsoeffekterna av ATMP är osäkra. Behandlingens långvariga effekt kan vara sämre än prognosticerat utfall, men det är osannolikt att den visar sig vara bättre än prognosticerat. Exempelvis kan patienten förväntas återgå till full hälsa efter behandling. Det är inte omöjligt att hälsan på lång sikt trots allt är lägre än full hälsa, men att behandlingen skulle leda till en hälsa bättre än full hälsa är inte möjligt.

Alternativa behandlingar kan tillkomma, priset på befintlig behandling kan sjunka. Priset på den befintliga behandlingen överskattas om beräkningen inte tar hänsyn till detta. Terapeutisk eller generisk konkurrens kan sänka priset på nuvarande behandling. Det är betydligt mindre sannolikt att den framtida konkurrensen skulle bli mindre, varför det kan hända att det förväntade priset på befintlig behandling överskattas.

²⁰ Kapitel 5 analyserar alternativet att vänta med beslut om ATMP, och kostnader exempelvis i försämrad hälsa förknippade med att vänta.

Innovationer kan leda till bättre hälsoeffekter för alternativet till ATMP-behandlingen. Att utvecklingen skulle leda till att jämförelsealternativet medför sämre hälsa i framtiden än idag är mycket osannolikt och skulle i så fall bero på exempelvis beslut baserade på felaktig information. Eftersom hälsovinsten av ATMP är skillnaden mellan hälsa med ATMP och hälsa givet att jämförelsealternativet väljs, leder också detta till en överskattning av den förväntade vinsten med att välja ATMP.

Om syftet med värderingen är att fastställa hur mycket samhället maximalt är villigt att betala för ATMP, måste utgångspunkten vara att utgå från förväntad hälsoeffekt, och förväntad kostnad för alternativ behandling. Därför måste prognoser för dessa tillföras till beräkningen av det inkrementella värdet av ATMP. Den följande diskussionen är mer översiktlig, med en mer formell modell i Appendix B.

4.1.2 Oersättlighet och osäkerhet

Hur svårt det är att värdera ATMP skiljer sig mycket från fall till fall. En viktig dimension i analysen är graden av *oersättlighet*, där vi med oersättlighet menar hur länge och med vilken hälsa en patient skulle leva utan ATMP-behandling. Frågan är inte densamma som den om den underliggande sjukdomen är allvarlig eller inte – om befintlig behandling möjliggör ett långt liv är inte behandling med ATMP lika viktig som den skulle vara om alternativ behandling saknades.

Ur informationssynpunkt är det lättare att hantera oersättningsbara ATMP, i och med att osäkerheten om kostnader och hälso nytta utan behandling är liten, i värsta fall med död inom relativt kort tid. När graden av oersättlighet å andra sidan är liten är den förväntade livslängden utan ATMP lång. Då måste vi värdera detta kontrafaktiska scenario över en lång tid. Hur skulle livet för dessa patienter se ut om tio år, utan ATMP-behandling?

I det mest extrema fallet, om ATMP enbart är kostnadsbesparande, är det avgörande hur kostnadsutvecklingen för det kontrafaktiska alternativet ser ut. Det är osannolikt att den faktiska besparingen över en mycket lång tidshorisont kommer att vara lika hög som den skulle bli om kostnaden för det nu befintliga jämförelsealternativet var oförändrat.

4.1.3 Diskontering

En viktig skillnad mellan ATMP och andra läkemedel är behandlingseffektens längd. En ATMP-behandling förmodas leverera hälsovinster för patienten långt in i framtiden. Effekten av vanliga läkemedel upphör ofta efter en tid, varför de ofta konsumeras kontinuerligt, i varje fall för kroniska sjukdomar. Den sjukvårdsbudget som vi har idag kan användas för behandlingar som ger oss omedelbara hälsovinster eller, alternativt, för behandlingar som framförallt ger upphov till hälsovinster i framtiden. För att kunna välja måste vi värdera hälsovinster vid olika tidpunkter mot varandra.

För att beskriva dessa värderingar kommer vi att använda en så kallad diskonteringsfaktor. Utgångspunkten i denna rapport är att ta diskonteringsfaktorn som given, för att fokusera på frågor om osäkerheter och värdering av dessa. Exempelvis kan den diskontering som TLV använder tas som utgångspunkt. Den analys som presenteras här är på inget sätt kopplat till valet av diskonteringsfaktor. Framtida osäkerheter introduceras på liknande sätt som en dödsrisk - de påverkar det förväntade värdet av behandlingen. I en separat

underlagsrapport till detta regeringsuppdrag diskuterar Johan Stennek ingående hur diskonteringsfaktorn (eller diskonteringsfaktorerna) kan bestämmas.

4.2 Osäkerhet om läkemedlets effekter

Värdebaserad prissättning bygger på idén att läkemedelsföretagen tillhandahåller evidens för läkemedlets värde. Det huvudsakliga utmaningen med värdering av ATMP är att företagen inte kan producera evidens med samma grad av säkerhet som för traditionella läkemedel. Orsaken är att de positiva effekterna förväntas bestå långt in i framtiden och att man ännu inte har erfarenhet av långsiktiga effekter.

De kliniska studierna ger knapphändig evidens

Jämfört med traditionella läkemedel uppstår det med ATMP nya osäkerheter när det gäller hur många QALY som faktiskt kommer att realiseras över tid, eftersom de studier som ligger till grund för godkännande ofta sträcker sig över kortare perioder än den förväntade återstående livslängden för patientgruppen. Studier kan exempelvis analysera osäkerhet kring hur väl olika patienter svarar på behandling, men eftersom de är relativt kortvariga kan de inte svara på huruvida det finns långsiktiga komplikationer eller andra faktorer som gör att överlevnad över tiden inte motsvarar den förväntade.

Det är viktigt att skilja på två typer av osäkerheter. En typ av osäkerhet är hur väl patienterna svarar på behandling, vissa patienter kanske inte alls uppvisar någon förbättring. En annan typ av osäkerhet är hur livskvaliteten utvecklar sig över tiden för de som svarar på behandling.²¹ Dessa osäkerheter är inte begränsade till ATMP, även andra läkemedel kan vara förknippade med osäkerheter vid introduktion, men för ATMP är problemen av större omfattning.

Hur osäkerheten om förväntad QALY för ATMP-behandling fångas i den hälsoekonomiska beräkningen

Med kontinuerlig behandling bär läkemedelsföretaget mycket av den risk som beror på osäkerhet i långsiktig hälsoeffekt. När ett läkemedel inte längre fungerar eller när patienten avlider upphör betalningen. För ATMP med betalning vid användning är betalningen oberoende av om läkemedlet visar sig vara mindre effektiv över tiden än förväntat.

För att korrekt värdera ett ATMP måste man bilda sig en uppfattning om vilka osäkerheter som finns med en behandling och inkludera dessa. Ett möjligt sätt att göra detta är att undersöka de osäkerheter som typiskt sett, det vill säga för nya läkemedel (nya ATMP) i allmänhet, påverkar välfärdsanalysen, såvida inte läkemedelsföretaget på ett övertygande sätt lyckas påvisa effektens storlek. En sådan *prior* baseras på generell erfarenhet om de osäkerheter som är kopplade till behandlingar.

En konservativ prior är att anta nolleffekt i avsaknad av statistiskt säkerställda studier, exempelvis för hur mycket av hälsoeffekten som kvarstår för tidsperioder bortom de som kunnat studeras. Detta är ju den hypotes som brukar ställas mot alternativhypotesen, att ett läkemedel har effekt, vid läkemedelsprövningar. En antagen nolleffekt bortom de tidsspann

²¹ Se Persson mfl (2019) för en utförlig diskussion av interaktionen mellan dessa osäkerheter.

som motsvarar längden på de effektstudier som ligger till grund för godkännande och prövning av hälsoekonomisk kostnadseffektivitet är dock en mycket konservativ princip. Se kapitel 7 för en vidare diskussion om hur hänsyn kan tas till denna typ av antaganden i betalningsmodellen.

4.3 Förväntad prisnedgång för jämförelsealternativet

Samhällets betalningsvilja definieras som priset på bästa jämförelsealternativ plus hälsoekonomisk värdering av den inkrementella hälsovinsten av det undersökta läkemedlet, jämfört med alternativet. Om priset på alternativet sjunker med ett visst antal kronor (punkt 3 på listan ovan) eller om hälsoeffekten stiger så mycket, att förbättringen värderas till lika många kronor vid oförändrat pris (punkt 2), minskar den maximala betalningsviljan för det undersökta läkemedlet med ett lika stort belopp. Om betalningen för ATMP görs med ett engångsbelopp kommer detta, sedan det betalats ut, att vara oberoende av om ett bättre alternativ skulle uppkomma. Hänsyn till förväntad framtida utveckling måste därför tas i förväg, när priset fastställs.

Även ifall hänsyn på detta sätt tas till förväntad framtida utveckling har ett ATMP med engångsbetalning fördelen, för tillverkaren, att *risk*en för framtida konkurrens minskar eller försvinner. När en patient är behandlad och betalningen har skett, är det vårdgivaren som står inför risken att vara låst vid ett sämre alternativ, exempelvis om utvecklingen av nya läkemedel för samma diagnos utvecklas snabbare än förväntat.

Hur ska framtida eventuell prisnedgång för jämförelsealternativet fångas i den hälsoekonomiska utvärderingen?

Det är relativt vanligt att nya läkemedel lanseras som alternativ till existerande redan under patenttiden. En möjlig utgångspunkt är att studera hur ofta innovationer historiskt sker inom patenttiden, det vill säga hur ofta terapeutisk konkurrens inträffar och vilken effekt det i så fall får. Även om detta sällan medför stora prisfall kan sådana förekomma; mer vanligt är att konkurrenterna tar marknadsandelar av det läkemedel som etablerar sig först.²²

Vår utgångspunkt är att nya läkemedel kommer att prissättas nära gränsen för samhällets betalningsvilja. Nya läkemedel med patent kan alltså inte *generellt* förväntas medföra en påtaglig överföring av överskott från läkemedelsbolag till samhället. Istället antar vi att det pristryck som finns framför allt visar sig efter patentutgång.

De reala läkemedelspriserna faller i genomsnitt med cirka 2 procent per år (e-Hälsomyndigheten; TLV). Ett korrekt konstruerat prisindex ska ta hänsyn såväl till pris- som kvalitetsutveckling. Vi bedömer att det uppmätta prisfallet i huvudsak beror på patentutgång och generisk konkurrens. En ansats vore därför att schablonmässigt anta att jämförelsealternativet faller med 2 procent per år. Emellertid har vi mer precis information än så: vi vet när patentet kommer att löpa ut för det aktuella ATMP och vi kan bilda oss en uppfattning om hur stort prisfall som då kan förväntas. Vi har i många fall också anledning att tro att inget betydande prisfall sker förrän vid patentutgång.

²² Mueller och Frenzel, 2015.

Vi kan därutöver ha specifik information om att jämförelsealternativet står inför patentutgång och förväntas falla i pris, eller att ett antal terapeutiska alternativ till aktuellt ATMP kommer att lanseras inom kort och att det kommer att gå att iscensätta en konkurrenssituation mellan dessa alternativ. Med specifik information av denna typ kan vi göra en annan prognos för prisutveckling av jämförelsealternativet.

4.4 Kostnader över tiden

De samhällsekonomiska kostnaderna för olika behandlingar kan också bli olika beroende på när betalningarna sker i tiden. Sjukvården har en årlig budget varje år. Denna budget bestäms av ett antal faktorer, och ger en budgetrestriktion för varje år. Som förklarat i avsnitt 3 väljer sjukvården de behandlingar som har ett ICER som ligger under tröskelvärdet. Introduceras en ny behandling, kan det hända att en behandling som har ett ICER lika med tröskelvärdet trängs undan. Alternativt finns det mindre utrymme för annan användning. För en mer utförlig diskussion, se Drummond mfl (2015, sid 108-112) och Claxton mfl. (2011).

Hur mycket det offentliga kan betala för en QALY på marginalen beror på ett antal faktorer utöver budgetutrymme. Teknologisk utveckling och konkurrens i sjukvårdssektorn jämfört med andra sektorer påverkar hur mycket hälsoinsatser kostar. Tröskelvärdet för ICER bör därmed variera, särskilt över långa tidsperioder som en patients livstid.

En beräkning av QALY med kvoten av nuvärden av kostnader och nytta innebär att man implicit antar ett konstant tröskelvärde över tiden, det vill säga att tröskelvärdet för kostnaden per QALY varje år är lika. Om betalningen sker omedelbart är detta inget problem, men om man ska jämföra en engångsbetalning nu med en kontinuerlig betalning över flera decennier kan det spela en stor roll.

För att illustrera hur olika betalningar kan påverka utfallet om sjukvårdsbudgeten blir "mer generös" i framtiden. Antag två behandlingar som bägge är över tröskelvärdet. En botande behandling bekostas nu, och har lite bättre ICER än en kontinuerlig som betalas lika mycket nu och i period 2. Om våra inkomster och sjukvårdsbudgeten stiger i period 2 kan valet av behandling 2 vara bättre, eftersom den vård som trängs undan i period 2 inte ger lika hög nytta per krona som den som kommer i period 1. En större sjukvårdsbudget i period 2 innebär att de insatser som trängs undan har lägre nytta än de som trängs undan i period 1.

Lorenzi et al, 2019, uppskattar att sjukvårdskostnaderna kommer att växa med 2,7 procent kommande decennier i OECD och ungefär samma i Sverige, vilket är mer än BNP. Ungefär hälften av tillväxten beror på inkomsteffekt, det vill säga att vi blir rikare, men andra halvan beror på att vi blir äldre och på Baumol-effekten. Kostnadspressen förväntas alltså i vissa avseenden bli *värre* kommande år. Slutsatsen är att även om det är sannolikt att den reala sjukvårdsbudgeten per capita kommer att vara större i framtiden och att vi därmed kommer att ha råd att höja ICER-gränsen, är denna slutsats inte helt självklar. Vissa bedömare varnar för att vi tvärt om kommer att tvingas sänka gränsvärdet.

Att hänsyn måste tas till omfördelning av vårdkostnader över tiden utöver diskontering av framtida QALY och kostnader har diskuterats länge (Drummond et al., 2015). I och med att

ATMP kan innebära att vård och kostnader för decennier framöver bekostas nu innebär att frågan blir högst relevant. Kan vi förvänta oss att tröskelvärdet för ICER kommer att vara konstant i decennier? Om inte, måste ICER-villkoret anpassas.

4.5 Hopp och oro

Persson m.fl. (2019) har framfört idén att ATMP ska värderas högre än vanliga läkemedel, eftersom de ger bot. Tanken är som följer. Säg att ett traditionellt läkemedel reducerar sannolikheten för att en patient ska dö till följd av en viss sjukdom från 100 procent till 1 procent. Denna effekt värderas enligt sedvanliga principer utifrån antal vunna QALY. Säg att ett ATMP reducerar sannolikheten från 100 procent till 0. Då bör ATMP-läkemedlet värderas upp mer än förhållandet 100/99. Anledningen är att den oro som patienten torde känna även när sannolikheten för död bara är 1 procent elimineras.

På motsvarande sätt kan man tänka sig att ett läkemedel som ger 1 procent chans till överlevnad är värt mer än en tiondel av ett läkemedel som botar med 10 procent chans. Orsaken är att redan det första ger patienten hopp.

Om man tar hänsyn till hopp och oro behöver förväntad nytta alltså inte vara linjär i sannolikheterna. En effekt som denna skulle kunna ha sin grund i de mekanismer Kahneman och Tversky beskrivit med sin *prospect theory*, enligt vilken individer tenderar att väga in utfall med låg sannolikhet med en högre vikt i sin nyttofunktion, än den som motsvarar sannolikheten för utfallet. En risk på 1 procent kommer enligt teorin att viktas högre, exempelvis till 5 procent, medan en risk på 99 procent vägs med en lägre vikt, exempelvis 95 procent.

Enligt teorin utgår också individerna från en referenspunkt, som i detta fall kan vara fullgod hälsa, när de värderar olika utfall. Utifrån referenspunkten antas individerna vara riskaversiva i förhållande till vinster men risksökande i förhållande till förluster. Med samma referenspunkt innebär detta att individer tenderar att föredra ett lotteri med en viss sannolikhet för ett väldigt negativt utfall och en viss sannolikhet för en oförändrad situation, framför ett säkert utfall med det förväntade värdet av lotteriet. Samtidigt föredrar samma individ det säkra utfallet motsvarande det förväntade värdet av ett lotteri som med en viss sannolikhet ger ett stort positivt utfall och med viss sannolikhet en oförändrad situation.

Beroende på vad som är referenspunkten kommer skillnaden mellan en låg dödsrisk och ingen dödsrisk att värderas högt eller lågt. Med full hälsa som referenspunkt är en låg dödsrisk just en risk, som värderas upp av tre skäl, enligt *prospect theory*. För det första värderas förluster högre än vinster, för det andra överviktas små sannolikheter och för det tredje är individer *risksökande* inför förlustrisker. Det senare innebär att värdefunktionen är brant.

Med säker död som referenspunkt kommer skillnaden mellan en låg dödsrisk och ingen dödsrisk att värderas mycket lägre. Fortfarande kommer låga sannolikheter att få en övervikt, men nu ses skillnaden mellan säker och nästan säker överlevnad som en vinst och inför vinster är individer *riskaversiva*, vilket i detta fall drar ner värderingen. Vinster värderas också, enligt teorin, lägre än lika stora förluster.

Huruvida det finns ett värde för hopp eller oro utöver dessa effekter är oklart. *Prospect theory* är dock empiriskt väldokumenterad och innebär att värdet som uppmäts för små

skillnader i dödsrisk kommer att vara känsligt för *framing*, det vill säga beroende på vad som uppfattas som referenspunkten. Det innebär att effekten är svårsmätbar, om den finns.

Ett relaterat problem är att det som uppges när en individ tillfrågas ofta inte stämmer särskilt bra överens med vad individen väljer och hur individen agerar. Problematiken med *stated and revealed preferences* innebär att vi måste välja vilken värdering samhället ska utgå ifrån i värdering av läkemedel: vad folk säger i en enkät eller hur de agerar till vardags.

Om värdet av bot betraktas som en konstant separabel effekt skulle det kunna motivera en något högre värdering. I förhållande till andra hälsoeffekter för ATMP återstår det dock att visa att värdet av bot ökar nyttan mer än marginellt. De hälsovinster som en behandling ger upphov till kan vara så stora att vinster som hopp och oro är av en mindre storleksordning.

Givet ovanstående problem och givet att utgångspunkten för den hälsoekonomiska värdering som TLV:s arbete bygger på är en systematisk och rationell värdering av hälsoutfall i linje med nyttoteorin, är vår slutsats att hopp och oro inte bör vägas in i bedömningen.

5 Inlåsningskostnader och behandlingsneutralitet

- *Vid värdering av ett nytt läkemedel jämförs kostnaden per QALY med den kostnad samhället maximalt är berett att betala. För traditionella läkemedel sjunker denna kostnad oftast över tiden på lång sikt. För ATMP med förskottsbetalning sker ingen motsvarande kostnadsnedgång.*
- *Samhället bör eftersträva en slags behandlingsneutralitet – en symmetri mellan behandlingsformer där ersättningen till den ena eller den andra typen av behandling bör vara lika, om man väger in de kostnads- och välfärdsskillnader som de medför.*
- *För att uppnå behandlingsneutralitet mellan ATMP och kontinuerliga behandlingar måste priset på ATMP ta hänsyn till patentutgång.*
- *Ett sätt att uppnå behandlingsneutralitet är att låta värderingen av läkemedel i huvudsak beakta enbart hälsovinster och kostnader före patentutgång.*
- *Att användningen av ATMP är oåterkallelig gör att det skapas ett optionsvärde för alternativa (traditionella) behandlingar. Särskilt om ICER är nära tröskelnivån och det är liten progression i sjukdomen utan ny behandling, kan värdet av att vänta vara högre än värdet att behandla nu.*

Den inlåsningskostnad som ATMP innebär medför en kostnad jämfört med traditionella läkemedel med kontinuerlig behandling. För kontinuerliga behandlingar brukar kostnaden ofta sjunka med tiden genom generisk eller terapeutisk konkurrens. I och med att ATMP leder till en inlåsningskostnad finns inte motsvarande utrymme för byte till annan behandling och, om ATMP ersätts med en engångsbetalning, inte heller för framtida prissänkningar. Vi har i föregående kapitel påpekat att om inte hänsyn tas till förväntad utveckling av priser och hälso nytta för alternativ till ATMP, kommer ATMP att övervärderas.

Vi fördjupar i detta kapitel diskussionen av principen om *behandlingsneutralitet*. Enligt principen bör den ersättning som samhället är villigt att betala vara lika om den förväntade netto nyttan är lika.

5.1 Behandlingsneutralitet

Bland andra Towse och Fenwick (2019) har anfört att botande och kontinuerliga behandlingar bör hanteras icke-diskriminerande. Tanken är att samhället bör eftersträva en slags *behandlingsneutralitet* – en symmetri mellan behandlingsformer där ersättningen till den ena eller den andra typen av behandling bör vara lika, om man väger in de kostnads- och välfärdsskillnader som de medför.

I Sverige och en del andra länder används värdebaserad prissättning för läkemedel, vilket vi beskrivit i kapitel 3. Som vi argumenterade i kapitel 4 talar tenderar priserna att falla över tid. Terapeutisk och generisk konkurrens leder till att den förväntade kostnaden per QALY i framtiden blir lägre än den är idag. Värdebaserad prissättning innebär att den kostnad samhället *maximalt* är villigt att betala är summan av kostnaden för jämförelsealternativet och värderingen av hälso nytta för den inkrementella hälsoeffekten. Att samhället är villigt att betala en kostnad innebär dock inte att den i normalfallet **faktiskt** betalar så mycket. Särskilt för sjukhusläkemedel har det ofta varit möjligt att få till stånd betydligt lägre priser, genom upphandling eller i förhandlingar. Vi återkommer till detta i kapitel 6.

Vi menar att behandlingsneutralitet bör gälla över tid. Ett ATMP bör värderas utifrån *förväntad* inkrementell nytta i förhållande till jämförelsealternativet, förväntad kostnad för jämförelsealternativet och förväntat tröskelvärde per QALY. Detta innebär att *värdet av* hälsonyttan i många fall kommer att falla, även om hälsonyttan är konstant.

Mer specifikt kan vi förvänta att värdet av hälsonyttan faller när patentet för ATMP löper ut, eftersom läkemedel efter patentutgång typiskt sett ersätts med det pris som etableras i konkurrens, vilket kan approximeras med produktionskostnaden. Värdet av hälsonyttan faller också när patentet för jämförelsealternativet löper ut, om detta ännu har patentskydd, liksom om det tillkommer nya mer effektiva läkemedel som är prissatta så att de ger större netto nytta för samhället än det gamla jämförelsealternativet. Därtill kommer diskontering, för det fall engångsbetalning används, liksom justering för osäkerhet om den långsiktiga effekten av ATMP.

I kapitel 4 introducerade vi begreppet oersättlighet. För ett ATMP som räddar patientens liv och där inget alternativ finns, är samhället villigt att betala för hela den återstående förväntade livstiden, även om den är längre än den återstående patenttiden. Vi menar dock att detta strider mot principen om behandlingsneutralitet. För ett traditionellt läkemedel erhåller läkemedelsföretaget intäkter i 15 år under patenttiden, om inte terapeutiska alternativ uppkommer. För ett ATMP med förskotts betalning innebär motsvarande pris säkra intäkter för patientens hela förväntade livstid. Denna ersättning kan potentiellt vara mycket större. Om förväntad livstid med ATMP behandling är 50 år, utgör ersättningen efter patentutgång om priser är oförändrade hela perioden ungefär hälften av den diskonterade framtida kostnaden. Kostnadsskillnaden för samhället kan också vara betydande, i den mån kostnaden per QALY sjunker på lång sikt.

Att kostnaderna för traditionella läkemedel tenderar att falla är viktig att beakta eftersom alla behandlingar ytterst konkurrerar om begränsade offentliga medel. Hur behandlingsneutralitet uppnås på bästa sätt bör utredas ordentligt. Inom ramen för denna rapport pekar vi på ett administrativt enkelt sätt att uppnå neutralitet. Om värdering av alla läkemedel tidsbegränsas till motsvarande patenttiden.

5.2 Värdering av ATMP över tiden

Principen om behandlingsneutralitet innebär att värderingen av ett ATMP, enligt ovan, tar hänsyn till patentutgångar och framtida konkurrens. Den oundvikliga konsekvensen är att värderingen av ATMP måste vara tidsberoende, det vill säga den kommer typiskt sett att sjunka över tiden.

Låt säga att ett nytt ATMP lanseras med en återstående patenttid på 15 år. Om patienternas förväntade överlevnad är längre än så måste priset på ATMP falla år för år, med start redan vid lanseringen. Om istället den förväntade överlevnaden är tio år börjar effekten av den kommande patentutgången påverka priset efter fem år.

Enligt ovan är vår uppfattning att graden av oersättlighet bör få avgöra om hänsyn även ska tas till *kommande* patentutgångar för jämförelsealternativet. Även för oersättliga ATMP kommer dock en patentutgång för jämförelsealternativet att slå igenom på värderingen av ATMP när patentet för alternativet väl löpt ut.

Situationen kan enklast beskrivas genom två exempel.

Exempel 1: Anta att ett ATMP lanseras med tio år kvar till patentutgång och att patientgruppen har en återstående överlevnadstid också på tio år. För enkelhets skull antas ingen diskontering samt att tillverkningskostnaden för jämförelsealternativet är noll. Jämförelsealternativet, som ger en hälsovinst som värderas till 300 000 kronor per år, har ett patent som löper ut om fem år. Den inkrementella hälsovinsten av ATMP värderas till 200 000 kronor per år, det vill säga utöver den hälsovinst som jämförelsealternativet ger, men det finns inga bestående effekter av att använda eller inte använda ATMP som vi behöver ta hänsyn till. Det vill säga ATMP-läkemedlet är i vår bemärkelse inte oersättligt. Tillverkningskostnaden för ATMP uppskattas till 1 miljon kronor eller 100 000 kronor per behandlad patient per år.

Priset för ATMP vid introduktion kan med dessa antaganden accepteras om det inte är högre än $5 \times 500\,000 + 5 \times 200\,000 = 3,5$ miljoner kronor.

Priset för ATMP ett år senare kan accepteras om det inte överstiger $4 \times 500\,000 + 5 \times 200\,000 + 100\,000 = 3\,100\,000$ kronor.

Priset för ATMP två år senare kan accepteras om det inte överstiger $3 \times 500\,000 + 5 \times 200\,000 + 2 \times 100\,000 = 2\,700\,000$ kronor.

Sex år efter lansering kan priset accepteras om det inte överstiger $4 \times 200\,000 + 6 \times 100\,000 = 1\,400\,000$ kronor och efter tio år kommer priset att stabiliseras vid en miljon kronor, givet att tillverkningskostnaden fortsatt är 1 miljon kronor och om inte terapeutisk konkurrens gjort att efterfrågan försvunnit.

Exempel 2: Vi gör samma antaganden som ovan, bortsett från att vi nu antar att ett nytt ATMP är oersättligt. Vi tar nu inte längre hänsyn till att patentet på jämförelsealternativet kommer att löpa ut innan det faktiskt löpt ut.

Priset för ATMP vid introduktion kan accepteras om det inte är högre än $10 \times 500\,000 = 5$ miljoner kronor.

Priset för ATMP ett år senare kan accepteras om det inte överstiger $9 \times 500\,000 + 100\,000 = 4,6$ Mkr.

Priset för ATMP två år senare kan accepteras om det inte överstiger $8 \times 500\,000 + 2 \times 100\,000 = 4,2$ Mkr.

Sex år efter lansering kan priset accepteras om det inte överstiger $4 \times 200\,000 + 6 \times 100\,000 = 1,4$ Mkr och efter tio år kommer priset att stabiliseras vid en miljon kronor.

För att behandlingsneutralitet ska gälla behöver alltså priset på ATMP följa en "glidbana", så länge priset baseras på en hälsoekonomisk värdering. När patentet löper ut kommer priset, som för alla läkemedel utan giltigt patent, att etableras genom en offentlig upphandling (för sjukhusförbrukning) eller möjligen i det auktionsförfarande som används för generiska läkemedel inom läkemedelsförmånen.

Vi understryker att ekonomisk teori enligt vår bedömning inte kan svara på om ovanstående är "rätt" sätt att prissätta ATMP, men vi bedömer att det är en rimlig princip.

5.3 Värdet av att vänta

När ett nytt läkemedel introduceras kan man ofta förvänta sig att andra läkemedel som löser samma problem är under utveckling. När läkemedlen är engångsbehandlingar uppstår då frågan om man ska välja läkemedel A som finns tillgängligt idag eller vänta till det att läkemedel B finns tillgängligt. Läkemedel B kan var billigare eller bättre, eller så kan konkurrensen mellan A och B leda till att ett lägre pris för A erbjuds. I detta läge uppstår det ett nytt villkor utöver ICER som måste vara uppfyllt. Även om det motiverat att välja A, kan det vara än bättre att vänta.

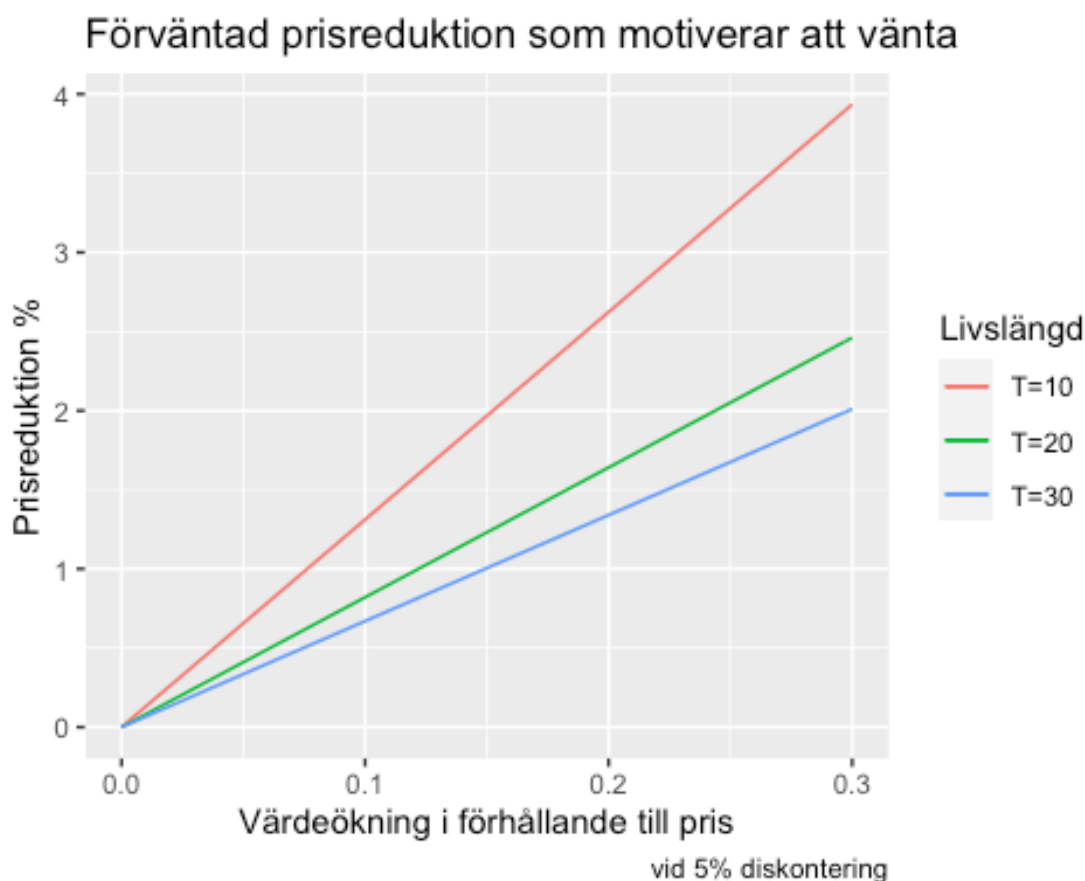
Att ett nytt läkemedel B förväntas komma ingår inte i en kalkyl av kostnad per QALY. Vid lansering av läkemedel A jämförs det med befintlig praxis eller med ingen behandling. Förväntade framtida alternativ bör dock beaktas, i och med att en botande behandling med A innebär att B inte kommer att användas.

Ett sätt att beakta värdet att vänta för alternativa behandlingar är att inkorporera dem i en prognos för det kontrafaktiska alternativet, i linje med analysen i kapitel 4. Det innebär att kommande alternativa innovationer ingår i värderingen av ATMP – i och med att förväntade innovationer gör det additionella värdet av ATMP lägre. Att göra en fullständig värdering kräver dock kunskap om sannolikheter för innovationer och värdet av dessa. Istället för att försöka värdera i förskott kan man fråga sig om man kan bedöma under vilka villkor det är bättre att vänta en att behandla nu. Eftersom det inte innebär en komplett värdering av alternativet till ATMP är det informationsmässigt mindre krävande.

Att vänta med en botande behandling kan medföra en välfärdsförlust i termer av lägre QALY för patienten. Om den alternativa behandlingen innebär lägre livskvalitet, innebär en uppskjuten botande behandling en QALY-förlust. Det kan också vara så att en försenad behandling leder till en försämrad livskvalitet under resten av livet, vilket också medför lägre QALY. Dessa förluster – som vi diskuterar i nästa avsnitt – kan dock vara motiverade om en väntan innebär att en bättre eller mindre kostsam behandling kan användas senare. Behandling "nu" kan också innebära att ny behandling krävs senare, om denna behandling är mer effektiv, vilket vi återkommer till i kapitel 6.

För icke livshotande sjukdomar kan det vara värt att skjuta upp behandling i väntan på billigare eller bättre behandlingar, även om dessa förbättringar är små. Det finns en risk att undervärdera värdet av att vänta, särskilt när värdet av behandlingen är utspritt över många år. En liten prissänkning eller kvalitetsförbättring som kan uppnås genom att vänta kan överväga kostnaden, i och med att förbättringen summeras över många år.

Välfärdsförlusten med att vänta beror också på hur mycket lägre ICER är än tröskelvärdet för godkännande. Om priset inklusive rabatter innebär att ICER knappt understiger tröskelvärdet, är nettokostnaden för att *inte* använda en ATMP låg. Då kan även små framtida prissänkningar eller små framtida förbättringar motivera att vänta. I figuren nedan illustreras vilken prisreduktion som skulle motivera att vänta med behandling, som funktion av värdeökningen (dvs *net benefit*) i förhållande till pris. När värdeökningen av den nya behandlingen är liten jämfört med befintlig krävs en mycket liten förväntad prisreduktion för att det ska vara bättre att vänta. Exempelvis gäller att om NB är tio procent av priset på läkemedlet kan det vara värt att vänta på en prisreduktion på mindre än en procent vid introduktionen om den förväntade livslängden med behandling är mer än 20 år, trots att en relativt hög diskonteringsränta använts. Modellen som figuren bygger på finns i Appendix C.



Om läkemedelsföretaget föreslår en prissättning som innebär att nyttoökningen är liten, kan även små framtida prissänkningar motivera att vänta. På motsvarande kan även en förväntan om små förbättringar i välbefinnande i kommande behandlingar motivera att vänta med behandling. Ett liknande resonemang gäller också om det finns en liten sannolikhet att en väsentligt bättre behandling lanseras i nästa period.²³

Resonemangen ovan gäller inte för behandlingar av livshotande sjukdomar eller där väntan leder till väsentligt lägre hälsa, då avsaknad av behandling påverkar alla tidsperioder framåt i tiden. Progression eller förväntade hälsoförbättringar går också att räkna med, se Appendix C.

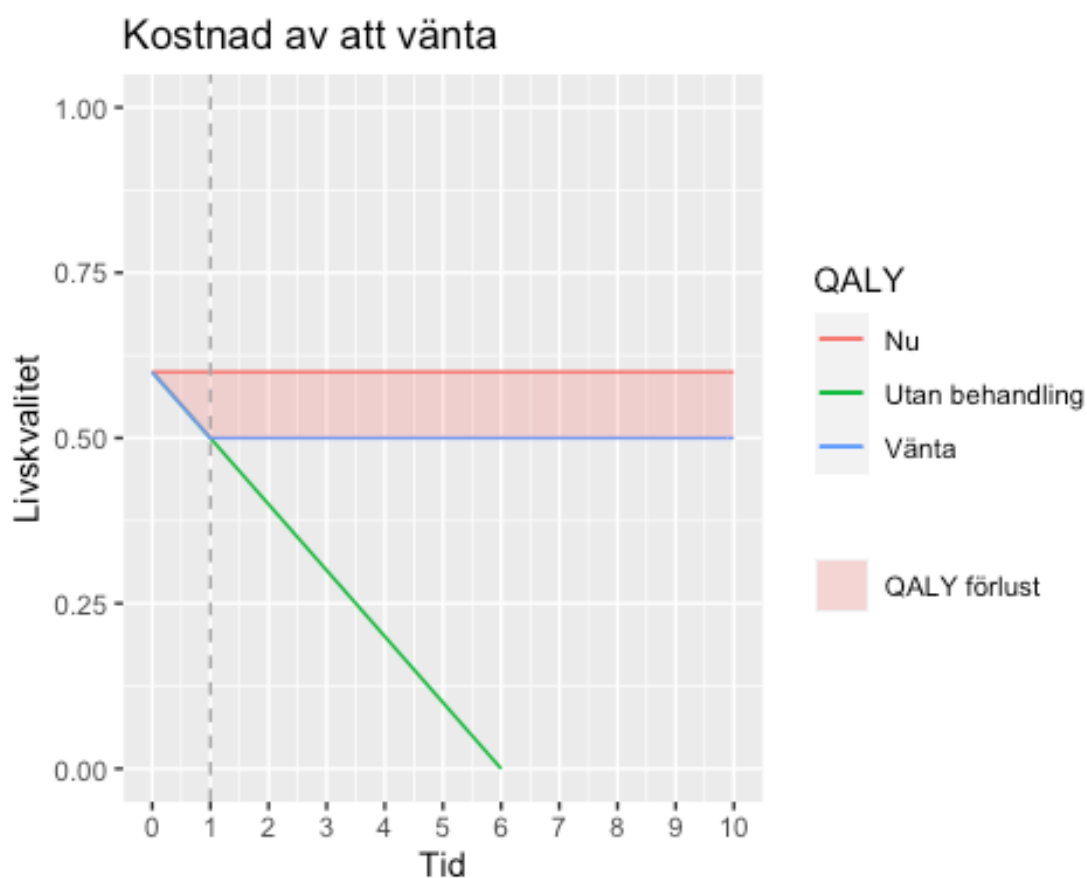
5.4 Kostnad av att vänta

Kostnaden av att vänta med en botande behandling beror på hur denna tid påverkar välfärden senare, när behandling genomförts. Om väntan innebär en permanent försämring av hälsotillståndet måste hänsyn tas till detta.

Figur X visar grafiskt hur en beräkning kan göras. Värdet av den permanenta försämringen av hälsotillståndet motsvarar hela det röda området i figuren. Här antas det att den botande behandlingen förlänger livet med 10 år om den påbörjas direkt, och ett år mindre

²³ Om vårdgivaren tar hänsyn till värdet att vänta, och detta är känt för A, skulle priset kunna sättas redan i första perioden så att det nya villkoret uppfylls. Hur realistiskt detta scenario är kan dock ifrågasättas, med tanke på upphandlingsregler, administrativ tröghet mm.

behandlingen dröjer ett år. Om fördröjningen inte har en bestående inverkan utgörs förlusten av det röda området till vänster om den vertikala streckade linjen.



Det skulle vara relativt enkelt kunna räkna på ett par konkreta fall, utifrån de QALY-beräkningar som gjorts, om man hade tillgång till dessa. Tyvärr är de dock inte offentliga.

Det kan också finnas det ett optionsvärde av en behandling som förlänger livet med ett år, eftersom det under detta år kan komma att utvecklas nya läkemedel som i bästa fall botar individen. Även i detta fall kan en förväntan om framtida behandlingar påverka värderingen av det livsuppehållande läkemedlet. På motsvarande sätt kan det finnas ett värde av att tillfälligt bromsa en sjukdoms progression som är kopplat till en förväntan om kommande botande läkemedel.

5.5 Kostnad för inkrementellt bättre behandlingar

I analysen ovan antog vi implicit att patienten *antingen* får behandling "nu" eller senare, det vill säga inte både och. Det kan dock visa sig att en patient som redan fått ett ATMP skulle ha nytta av en ännu bättre behandling som utvecklas senare. Samhället kan då tvingas betala ett högt pris för framtida inkrementella förbättringar. Anta att ett ATMP förbättrar hälsan men inte återställer den. Anta vidare att det utvecklas nya läkemedel med en bättre effekt, som kan användas av redan behandlade patienter. Kommer samhället att begränsa tillgången till det nya bättre läkemedlet enbart till de patienter som inte fått det äldre ATMP? Om hälso nytta beräknas för obehandlade patientpopulationer kan ett högt pris

sättas för nya läkemedel. Kanske blir det svårt att upprätthålla en princip som innebär att patienter som fått ATMP inte får tillgång till det nya läkemedlet, även om ICER för denna patientgrupp är för hög.

Om vårdgivaren enbart betalar för den förbättring som den nya behandlingen ger upphov till, det inkrementella värdet, blir det en form av indikationsbaserad prissättning. Situationen innebär inte något problem. Dock är det inte givet att läkemedelstillverkaren kommer att önska en sådan differentierad prissättning.

5.6 Optionsvärde och att vänta

I resonemangen ovan har vi i huvudsak betraktat utvecklingen som deterministisk. Vi har visat på ett antal effekter som riskerar att övervärdera ATMP i förhållande till traditionella läkemedel. Emellertid kan det uppstå ytterligare fördelar för traditionella läkemedel genom att kontinuerlig behandling inte på samma sätt är oåterkallelig. Det kan finnas ett optionsvärde i att ha kvar handlingsfriheten.

Precis som i teorin för reala optioner (Dixit och Pindyck, 1994) tillkommer ett nytt villkor: att undersöka om det inkrementella värdet av att vänta är än större än att påbörja behandling nu. Detta är förstås bara möjligt om sjukdomen inte är dödlig eller om det finns alternativa läkemedel som erbjuder rimligt god behandling. Det vill säga, det finns inget optionsvärde att ta hänsyn till om alternativet är att patienten dör utan ATMP.

Det kan vara bättre att vänta med ATMP tills osäkerheten kring de långsiktiga effekterna har minskat, det kan vara bättre att vänta med ATMP tills priset fallit, exempelvis efter patentutgång. Det kan vara bättre att vänta och se om pågående FoU-projekt leder till att alternativ erbjuds som ger större nettonytta för samhället.

Att hänsyn tas till värdet att vänta behöver inte innebära att det nödvändigtvis leder till fördröjd behandling. Om prissättningen på ATMP anpassas till framtida alternativ leder till att kostnaden för vad samhället är villig att betala nu, kan det leda till att priset i dagsläget blir lägre. Det kan vara lönsamt för tillverkaren av ATMP att sänka priset för att undvika att köparen väntar på andra alternativ.

Vi illustrerar de olika sätten ett optionsvärde kan göra att det är lönsamt att vänta.

Exempel 1: En viss indikation kan behandlas med ett ATMP vi kallar för A, med ett traditionellt läkemedel B eller – kanske – med ett läkemedel C som lanseras nästa år. A och B ger lika bra hälsoeffekt, en hälsoeffekt som värderas till 200 000 kr per år i tio år, vilket är förväntad överlevnad i patientpopulationen. För enkelhets skull gör vi ingen diskontering. B kostar 100 000 kr per år och ger alltså en nettonytta på totalt 1 Mkr. Med 50 % sannolikhet kommer C att ge en hälsoeffekt som värderas till 300 000 kr per år och kosta 100 000 kr per år. Med 50 % kommer utvecklingen av C att misslyckas.

Förväntad nytta av C för befintlig patientgrupp är $0,5 \times (300\,000 - 100\,000) \times 9 = 900\,000$. C ser alltså ut som ett sämre alternativ än B och A kan till synes prissättas upp till 1 Mkr och ge samma nytta som B samt bättre nytta än C. Men det finns ett optionsvärde vi bör ta hänsyn till. Om utvecklingen av C lyckas väljer vi detta alternativ; om inte håller vi fast vid B. Detta ger en total förväntad nettonytta på $100\,000 + 0,5 \times 9 \times 100\,000 + 0,5 \times 9 \times 200\,000$ kr = 1 450 000 kr. Läkemedel A kan inte

säljas dyrare än 550 000 kr om netto nyttan ska bli lika hög. Läkemedlet C erbjuder ett optionsvärde för det offentliga på 450 000 kr, vilket utgör en alternativkostnad för att välja A direkt.

Längre bindningstider brukar ofta vara kopplade till lägre pris. Exempelvis brukar abonnemang med längre bindningstid ha lägre pris per tidsenhet. I och med att köparen begränsar sin valfrihet krävs ett lägre pris. Omvänt är det en fördel för säljaren att försäkra sig om en stabil efterfrågan och längre planeringshorisont.

I exemplet ovan är det osäkerhet om ett tredje alternativ som skapar ett optionsvärde som ökar lönsamheten för alternativet "vänta". Ett annat scenario är att det finns osäkerhet om effekten på ATMP, men att denna osäkerhet kommer att minska eller försvinna i framtiden. I föregående avsnitt har vi diskuterat detta scenario utifrån perspektivet att effekten på ATMP kan upphöra och att vi måste ta hänsyn till detta i prissättningen. Men finns det utöver detta ett optionsvärde?

Exempel 2: En viss indikation kan behandlas med ett ATMP kallat A. Vi vet att A fungerar i minst ett år. Med 50 % sannolikhet kommer det att fungera i tio år, men med lika stor sannolikhet bara fungera ett år; nästa år får vi reda på vilket som gäller. Vi kan istället använda det traditionella läkemedlet B som kostar 150 000 kr per år och ger hälso nytta motsvarande 200 000 kr per år. A ger samma hälso nytta – om det fungerar. Tillverkaren av A föreslår priset 600 000 kr. Förväntat netto nytta av A är då $200\,000 + 0,5 \times 9 \times 200\,000 - 600\,000$ kr = 0,5 Mkr, vilket ger samma förväntade netto nytta som att välja B alla tio år.

Om priset är detsamma nästa år bör emellertid erbjudandet inte accepteras. Att använda B det första året och bara välja A om det visar sig fungera på lång sikt ger oss netto nyttan $100\,000 + 0,5 \times (9 \times 200\,000 - 600\,000) + 0,5 \times 9 \times 50\,000$ kr = 925 000 kr. Friheten att välja A först efter att osäkerheten försvunnit skapar ett optionsvärde på 425 000 kr.

Exemplet bygger dock på att priset på A är detsamma utan som med osäkerhet. I detta fall kan tillverkaren av A vilja höja priset till närmare 1,5 Mkr. (Givet en förväntad återstående livslängd på tio år för hela patientpopulationen.) Om priset höjs i motsvarande mån som osäkerheten reduceras finns inget optionsvärde i att vänta.

6 Konkurrens och incitament för att investera i FoU

- *Vid upphandling av ATMP bör kostnaden för jämförelsebehandlingen, om sådan finns, inte tas som given. I TLV:s analys bör olika rabattnivåer för befintlig behandling ingå, för att underlätta terapeutisk konkurrens.*
- *Avsaknad av behandlingsneutralitet kan leda till snedvridning i FoU, med större incitament att utveckla ATMP relativt andra läkemedel.*

Värdering av nya läkemedel med kostnadseffektivitetsanalys och upphandling är i många avseenden snarlik annan offentlig upphandling. Infrastrukturinvesteringar som exempelvis vägbyggen utvärderas med en kostnadsintäktanalys och upphandlas. Det finns en avgörande skillnad mellan läkemedel och merparten av annan upphandling: det finns bara en anbudsgivare som dessutom är ett legalt monopol i ca femton år.

För att renodla resonemanget, antar vi att läkemedelsföretagen vinstmaximerar med full insyn i den offentliga värderingen. Avsaknaden av konkurrens gör att värderingen, vars syfte är att beräkna hur mycket samhället maximalt är villigt att betala, också blir grunden för prissättningen. I och med att läkemedelsföretaget har god insikt i vad samhället är beredd att betala, kommer det bli vad samhället betalar. I och med att producenten har full insyn i betalningsviljan, kommer all välfärdsökning som det nya läkemedlet ger upphov till att tillfalla företaget, åtminstone initialt.²⁴

Även om företaget har patent uppstår över tiden konkurrens på olika sätt, exempelvis genom terapeutisk konkurrens med ett biosimilar-läkemedel. Det innebär att det priset sjunker över tiden, och att fördelningen av överskottet förändras.

Även om läkemedelsföretaget har patent innebär det inte alltid att priset i praktiken bestäms av tröskelvärdet. Budgetbegränsningar kan göra att priset blir lägre. Som diskuterats i Kapitel 3, om läkemedlet för en region innebär en stor kostnadsökning och läkemedelsbudgeten är relativt fix, kan det göra att priset blir lägre. Givet att det finns en trovärdig budgetbegränsning som inte går att ändra på kort sikt, måste priset för det nya läkemedlet sänkas för att rymmas inom budgeten. För läkemedelsföretaget skulle innovativa betalningsmodeller kunna vara ett sätt att försöka mildra budgetrestriktionen för att på så sätt kunna sätta ett högre pris.

För läkemedel som berör stora patientgrupper med en stor hälsoeffekt finns det skäl att tro att prissättningen kommer att bestämmas genom förhandling. Ett aktuellt exempel på detta är EU:s upphandling av Coronavacciner. Det innebär att principer för värdering och upphandling av ATMP kanske kan fokusera på ovanligare sjukdomar.

En tolkning av detta är att läkemedelsbolaget får tillgodogöra sig i stort sett hela den hälsovinst som uppstår relativt *best practice*, men bara under den återstående patenttiden. Sjukvården kan också välja en annan behandling, om sådan skulle uppstå, eller avsluta behandling om den inte längre bedöms vara prisvärd exempelvis genom att påverka läkarnas förskrivningsmönster, till exempel genom "Kloka listan". Om motsvarande ansats tillämpas på ATMP behöver hänsyn tas till osäkerhet både angående långsiktig effekt på hälsan och angående bästa alternativs hälsoeffekter och kostnad.

²⁴ Det innebär att läkemedelsföretaget får en större del av överskottet än vad en monopolist som inte känner till varje individs värdering får. Om individer har olika värdering och monopolisten bara kan sätta ett pris för alla kommer de som har en värdering högre än detta pris få ett överskott.

För prissättningen av läkemedel för sjukhusförbrukningen används hälsoekonomiskt värde, relativt *best practice*, som ett tak. Regionerna kan dock ibland förhandla ner priset i förhållande till taket, även för läkemedel med giltigt patent, i synnerhet om det finns flera läkemedel som kan användas för samma indikation. Vi uppfattar dock att "taket", det vill säga tröskelvärdet för ICER, på motsvarande sätt som för läkemedelsförmånen motsvarar hela det inkrementella hälsoekonomiska värdet plus kostnaden för *best practice*.

6.1 Terapeutisk konkurrens

Vid bedömningen av en ny ATMP-behandling är det ofta i jämförelse med en befintlig behandling. Det är ett exempel på terapeutisk konkurrens, där valet står mellan alternativa läkemedel och priset för respektive ingår i detta val. Värdet av den nya behandlingen baseras delvis på den befintliga: en del av värdet av den nya behandlingen är att kostnaden för den befintliga upphör. Denna kostnad kan vara betydande och värdet av nya behandlingen kan då till stor del bestå i sänkta vårdkostnaderna. Det är dock inte klart att priset för den befintliga behandlingen bör betraktas som givet vid bedömningen av det nya läkemedlet. Den terapeutiska konkurrens som uppstår då ATMP-behandlingen introduceras kan ge incitament till producenten av den befintliga behandlingen att sänka priset. Det sänkta priset medför att ICER för ATMP-behandlingen blir högre.

För att illustrera vikten av att inte ta priset för jämförelsealternativet som given, antag att det uppkommer en rad nya alternativa läkemedel över tiden med exakt samma QALY-värdering.²⁵ När det första alternativa läkemedlet introduceras sätts priset marginellt lägre än befintlig behandling. Trots att nya alternativa behandlingar tillkommer, uppstår ingen prissänkande konkurrens.

6.1.1 Metoder för att stimulera till terapeutisk konkurrens med ATMP

Det finns exempel på att introduktionen av en ATMP-behandling föranlett en dialog med producenten av den befintliga behandlingen. Det skulle dock vara av värde att denna process formaliseras, för att underlätta att en konkurrens uppstår. Utformningen av processen är dock inte helt fri. Regionens inköp av läkemedel lyder under upphandlingslagstiftningen, vilket sätter begränsningar på interaktionen mellan upphandlande region och anbudsgivare.

I nuvarande system gör TLV en bedömning av det terapeutiska värdet av ett nytt läkemedel. Utifrån det behandlingspris som företaget bestämt beräknas sedan, vid olika rabattnivåer, det ICER den nya behandlingen ger upphov till. Utifrån en tabell med rabattsatser och motsvarande ICER bedömer sedan NT-rådet om behandlingen har ett ICER som understiger det relevanta tröskelvärdet, och därmed bör rekommenderas som behandling. I analysen tas priset på rådande behandling som given.

Upphandling av läkemedel görs av regionerna utifrån NT-rådets rekommendationer. Det innebär att ett anbud från läkemedelsföretaget accepteras om behandlingskostnaden inte överstiger NT-rådets rekommendation.

Om den förhandling som sker mellan NT-rådet och läkemedelsföretag leder till en överenskommelse, ligger den ICER som rabatten innebär oftast nära tröskelvärdet. Det

²⁵ Eftersom de har samma QALY, kan de inte jämföras med ICER i och med att en kvot där man delar med noll är obestämd. Alternativen går dock att jämföra med NB.

innebär att merparten av överskottet av överenskommelsen tillfaller producenten. Eftersom den värdering och beslutsprocess som det offentliga tillämpar är känd av producenten är detta ett exempel på första gradens prisdiskriminering.

Det underlag som TLV utarbetar bör ta hänsyn till möjliga prisrabatter för befintlig behandling och ATMP, och innehålla beräkningsstöd för en jämförelse mellan behandlingar vid olika rabattsatser för olika tröskelvärden för ATMP. Det kan göras på liknande sätt som nuvarande underlag. I nuvarande analysordning ingår en tabell med ICER för ATMP som funktion av rabattsats, givet rådande pris på befintlig behandling. På motsvarande sätt kan man beräkna motsvarande ICER för ATMP om befintlig behandling ger rabatt.

En upphandling med terapeutisk konkurrens mellan befintlig behandling och ATMP skulle också kunna utformas med utvärdering av pris och kvalitet. Kvaliteten bestäms som QALY för respektive behandling. Viktning av pris och kvalitet görs genom en Net Benefit-analys, där kvalitet viktas med det relevanta tröskelvärdet för ICER för att få ett monetärt värde som inkluderar pris och kvalitet. Det anbud som har högst NB accepteras.

En tydligare inköpsprocess vid terapeutisk konkurrens kan också öka incitamenten att introducera nya läkemedel. Tydliga konkurrensregler minskar osäkerheten för företagen i att ta fram nya läkemedel, eller att hitta nya användningsområden för befintliga.

6.2 Investeringsincitament och konkurrensneutralitet

Traditionella läkemedel är enkla att kopiera. Konkurrerande bolag är dock under patenttiden förhindrade att göra det, vilket upprätthåller incitamenten att investera i forskning och utveckling. Under patentets giltighetstid är patentinnehavaren i princip fri att utnyttja sin ensamställning genom att ta ut höga priser. När patentet löpt ut är det däremot - också i princip - fritt att sälja generiska läkemedel, vilket brukar leda till en dramatisk prisnedgång.

Patentlängden kan ses som resultatet av en avvägning av motverkande effekter. En lång patenttid ger starka incitament för forskning och utveckling, men medför också att priserna förblir höga en lång tid. Långa patenttider kan också motverka utvecklingen av nya produkter som inkluderar den gamla produkten som en komponent. Detta är mindre viktigt i läkemedelsindustrin än i vissa andra branscher, men kan likväl vara en faktor om det är aktuellt att ta fram kombinationsläkemedel.

Nackdelarna med höga priser är påtaglig för nya läkemedel. I den utsträckning efterfrågan är priskänslig tränger höga priser undan efterfrågan, vilket leder till att konsumtionen av det patentskyddade läkemedlet blir alltför låg. I den utsträckning efterfrågan *inte* är priskänslig, till exempel för att de läkare som beslutar om läkemedelsanvändning inte i tillräckligt hög utsträckning tar hänsyn till kostnader, uppstår en annan typ av problem. Då överförs onödigt stora resurser från skattebetalare (eller från ett försäkringskollektiv) till tillverkarna och incitamenten att utveckla nya produkter blir eventuellt alltför starka.

För att hantera detta problem har flertalet länder, med USA som viktigt undantag, någon form av de facto prisreglering av patentskyddade läkemedel, vilket vi diskuterat bland annat i kapitel 3 och vilket förstås verkar återhållande på investeringsincitamenten.²⁶

Slutsatsen av ovanstående resonemang är att det inte är självklart att regelverket för prissättning av traditionella läkemedel är optimalt utformat. I själva verket finns det anledning att tro att prissättning och effektiv patentlängd för alla eller flertalet läkemedel avviker från vad som vore optimalt.

Vår utgångspunkt i den fortsatta analysen är dock att det nuvarande systemet på ett acceptabelt sätt balanserar de olika effekterna mot varandra och att det därför är önskvärt att incitamenten för att investera i utveckling av nya produkter är neutrala mellan ATMP och traditionella läkemedel, samt att samhället uttrycker samma betalningsvilja för ATMP som för ett traditionellt läkemedel som ger samma hälso nytta.

6.2.1 Investeringsincitament

Ett företag som investerar i FoU för ett traditionellt läkemedel kan, om projektet lyckas, räkna med en effektiv patenntid på maximalt 15 år.²⁷ Därefter försvinner ofta i varje fall merparten av vinsten genom konkurrens från generikatillverkare, som enkelt uttryckt driver ner priset till i nivå med kostnaderna. Men även under patenntiden är det vanligt att det uppstår konkurrens från andra nya läkemedel med patent. I den konkurrensen, så kallad terapeutisk konkurrens, är priset normalt sett inte det huvudsakliga konkurrensmedlet, utan istället den medicinska effekten. I många fall kommer det första läkemedlets marknadsandel successivt begränsas genom en ökande terapeutisk konkurrens.

För ett företag som utvecklar ATMP gäller samma regler som för övriga läkemedel, men konsekvenserna blir annorlunda om det nya läkemedlet ersätter en långvarig behandling med ett traditionellt läkemedel.²⁸ En patient som behandlats med en ATMP antas inte ha något ytterligare behov av läkemedel för den aktuella sjukdomen och bidrar därför inte till marknadspotentialen för ett konkurrerande läkemedel inom samma terapeutiska grupp. Vidare antas patienten inte ha behov av generiska alternativ, eftersom effekten antas vara livslång.

Det senare förhållandet kan hanteras genom att hälsovinster efter patentets utgång värderas lägre, nämligen till förväntad produktionskostnad. Betalningsmodellen för ATMP kan ge tillverkaren en ström av intäkter, så länge patienten lever och effekten består. Alternativt kan betalningsmodellen bygga på en engångsersättning för varje ny patient som behandlas med ATMP-läkemedlet, men där ersättningen krymper successivt, i takt med att patentutgången närmar sig. Vi diskuterar betalningsmodeller mer utförligt i nästa avsnitt.

Det förra förhållandet, att marknadspotentialen blir mindre för terapeutiska alternativ, gör det relativt sett mer attraktivt att vara först med en ny kategori av läkemedel. Den första tillverkaren kommer att kunna konkurrera om alla patienter som har en viss sjukdom. En ATMP eller ett traditionellt läkemedel som kommer senare till marknaden kommer potentiellt bara att kunna konkurrera om nytillkommande patienter. Effekten är som störst

²⁶ I USA används andra mekanismer för att begränsa läkemedelskostnader, såsom högre egenavgifter för dyra läkemedel (eller att de inte alls täcks av sjukförsäkringar?).

²⁷ Scott Morton och Boller, 2017 uppskattar medeltiden med effektivt patent i USA till 12 år,

²⁸ Scott Morton och Boller, 2017, menar dock att biologiska läkemedel inte möter effektiv konkurrens ens efter att patentet löpt ut.

för sjukdomar där man överlever länge även utan behandling, och det således finns många att behandla initialt.²⁹

I Appendix D visar vi att i och med att ett ATMP kan få en större del av överskottet genom engångsbetalningar finns starkare investeringsincitament för ATMP än för lika lovande traditionella läkemedel. Det innebär att investeringar allt för mycket kommer att inriktas på ATMP jämfört med traditionella läkemedel.

²⁹ Se också diskussionen i avsnitt 8.3 om initiala kostnader för det offentliga.

7 Betalningsmodeller

- *Att värdera hälso nyttan av ATMP för hela patientens förväntade livstid är mycket mer komplicerat än att värdera traditionella läkemedel.*
- *Samhället måste fastställa sannolikheter för möjliga utfall även om inte specifik information finns för den aktuella behandlingen.*
- *Med kontinuerlig betalning betingad på utfall är det lättare att hantera osäkerheter, utan att behöva bedöma dem i förväg. Vår bedömning är att detta sammantaget är att föredra ur ekonomisk synpunkt.*
- *Modeller med kontinuerlig betalning kan dock vara svåra att få till stånd, på grund av juridiska och administrativa problem.*
- *För att ge tillverkarna incitament att såväl investera i informationsinsamling som att överlämna all relevant information till myndigheterna, föreslår vi att om engångsbetalningar ändå används, ska de utöver de aspekter vi beskrivit i tidigare kapitel, även utformas så att det blir "konservativa"*

Som vi såg i föregående kapitel medför den skillnad som finns mellan traditionella läkemedel och ATMP med förskottsbetalning att den hälsoekonomiska värderingen blir betydligt mer komplex. Genom den inlåsning som ATMP innebär tas beslut om vård och om ersättning för denna "nu", snarare än vid många tidpunkter över tiden. Istället för att ta ställning till framtida utfall eller behandlingar allt eftersom, måste dessa förutses vid värderingen initialt.

Det ställer mycket stora krav på TLV:s analyskapacitet, särskilt i de fall där hälsoförlusten med annan behandling är liten. Vad ska vi tro om hälsan för patienter som inte behandlas med ATMP ett eller flera decennier i framtiden? Vad är sannolikheten att läkemedel som minskar biverkningar av nuvarande behandling introduceras om fem eller tjugo år? Med ett traditionellt läkemedel behöver vi normalt sett inte ställa dessa frågor. Om nya bättre alternativ uppstår byter vi behandling. Samma sak om patentutgångar eller ökad terapeutisk konkurrens för alternativ förändrar relativpriserna till nackdel för det läkemedel som först användes.

I kapitel 4 visar vi att engångsbetalning av ATMP leder till en övervärdering om inte hänsyn tas till framtida valmöjligheter. Vi behöver ta hänsyn till förväntad prisnedgång för jämförelsealternativet, till möjligheten att terapeutisk konkurrens driver ned priset liksom till risken för att ATMP inte fungerar så bra som förväntat. Dessutom måste framtida nytta diskonteras. Vidare bör priset på ATMP sjunka i takt med att patentutgång närmar sig, enligt föregående kapitel, för att följa principen om behandlingsneutralitet.

I och med att det kan vara svårt att göra prognoser över framtida utveckling och med hänsyn till att den hälsoekonomiska analysen riskerar att bli mycket komplex, bör möjligheten att ersätta engångsbetalning med någon annan betalningsmodell undersökas. Om ersättningen för ATMP görs lik ersättningen för traditionella läkemedel, blir informationsproblemen mindre. Vi behöver inte ta ställning till möjliga utfall i framtiden om framtida ersättning betingas på utfall, på motsvarande sätt som för ett traditionellt läkemedel.

Från ett ekonomiskt perspektiv talar det mesta för kontrakt med kontinuerlig betalning, betingad på utfall. Det som talar för engångsbetalningar är framför allt juridiska hinder och kontraktuella svårigheter, men också ett ekonomiskt argument:

- Det kan finnas juridiska hinder mot att betalning för ATMP sker över tid, i takt med att hälsovinster realiseras.
- Betalning över tid bör betingas på det medicinska utfallet, men detta kan vara svårt eller kostsamt att mäta, utöver enkla kriterier som till exempel att en behandlad patienten fortsatt är vid liv.
- Samhället har troligen lägre diskonteringsränta än läkemedelsföretagen, vilket gör att det i princip finns en kostnadsänkande effekt av engångsbetalningar.

Eftersom vi ändå menar att de ekonomiska argumenten sammantaget talar för en betingad kontinuerlig betalningsmodell diskuterar vi i det här kapitlet diskuterar hur alternativa betalningsmodeller skulle kunna utformas.

Vi inleder med en redogörelse för hur liknande problem, som de som gäller för prissättning av ATMP, har lösts på andra områden. Därefter diskuterar vi en del av de juridiska och administrativa problem som kan göra alternativa betalningsmodeller svåra att implementera i praktiken.

Om den ekonomiska analysen talar för löpande betalningar finns eventuellt möjlighet att hitta en betalningsmodell som fungerar juridiskt och som inte är alltför olik den ur ekonomisk synpunkt önskvärda modellen. Alternativt kan lagstiftning ändra de juridiska förutsättningarna.

För ATMP-behandlingar med små patientgrupper kan det hända att kontrakt baserat på utfall i svensk sjukvård är otillräckligt. För att bedöma den långsiktiga effektiviteten av en behandling kan det vara nödvändigt att studera behandlingen internationellt.

7.1 Betalningsmodeller – en översikt

Betalningar för varor, tjänster och tillgångar av olika slag kan struktureras på olika sätt, beroende bland annat på vilken information köparen respektive säljaren har om produkten, på beloppets storlek och på produktens natur.

Betalning sker normalt vid leverans om inte något annat avtalats. Det gäller både dagligvaror och kapitalvaror. Den främsta orsaken är sannolikt att båda parter har en hållhake på varandra fram till dess de har fullgjort sina skyldigheter. Köparen betalar inte förrän varan levererats. Säljaren levererar inte förrän betalningen erlagts.

För kapitalvaror är det inte ovanligt att betalningen delas upp. Det kan ske genom att varan hyrs ut istället för att säljas, till exempel genom ett leasingavtal. Det kan även ske genom att varan säljs men samtidigt ställs som säkerhet för framtida betalning. Dessa konstruktioner är dock inte aktuella när det gäller läkemedel eftersom behandlingen inte kan användas som säkerhet.

Ytterligare en möjlighet är förstås att säljaren lånar ut (eller förmedlar ett sådant lån från en tredje part) en del av beloppet mot andra säkerheter eller att man av andra anledningar har förtroende för att köparen kommer att betala, genom ett avbetalningsköp.

Dessa konstruktioner används ofta då köparen har en sämre tillgång till kapitalmarknaderna än säljaren. När det gäller läkemedel torde det vara tvärt om. Den svenska statens låneränta är sannolikt lägre än de flesta läkemedelsföretagens låneränta.

När anledningen till att full betalning inte görs omedelbart istället har att göra med informationsproblem används andra betalningsmodeller. Vid företagsköp är det inte ovanligt att köparen och säljaren överenskommer om ett tilläggsbelopp som är beroende av exempelvis vinst eller försäljning. Den rörliga delen av betalningen kan vara stor eller liten. Betalningen kan komma efter en kortare eller längre tid, men vanligen efter ett par år, och då ofta betingat på vinsten. Ett alternativ är att säljaren under en relativt lång period, kanske uppåt 20 år eller möjligen "för evigt", får en royalty eller motsvarande, det vill säga en procentuell andel av köparens försäljning inom den köpta verksamheten.³⁰

Immateriella rättigheter är ytterligare ett sätt att betala över tid för en prestation som utförts mycket tidigare. Patenträtt är den immaterialrätt som används för läkemedel. I det fallet binds ofta betalningen till en producerad vara och betalningen för patenträtten görs i samband med att varan (läkemedlet) köps. Vid sidan patenträtt finns även upphovsrätt, varumärkesrätt och mönsterrätt. I många fall går det inte tydligt att knyta betalningen för den immateriella rättigheten till en vara som säljs. Då knyts betalningen istället till användningen, genom royalty.

PRV beskriver på sin hemsida betalning för royalty på följande sätt:

"[Royalty utgår för] rätten att trycka och distribuera en bok, spela musik på en klubb, använda ett patent, mönster eller ett varumärke. Storleken på den royalty som betalas ut behöver inte vara ett fast belopp, utan kan baseras på försäljning, antal spelningar eller tillverkade produkter. Vid försäljning beräknas ofta royalty som en procentsats på uppnådd intäkt av försäljning, exempelvis vid franchising. Ett alternativ till ersättning i form av royalty är att licenstagaren betalar en engångsersättning, till exempel om man köper rätten att använda en licens av en mjukvara. Det är vanligt att licenstagaren initialt betalar en engångssumma, för att därefter betala royalty baserat på försäljning. Storlek på royalty, på vilket sätt den ska betalas och andra villkor regleras i ett licensavtal mellan den som äger rättigheterna (licensgivaren) och den som vill använda rättigheterna (licenstagaren)."³¹

Betalningsströmmen kan alltså i många fall anpassas över tid på ett ändamålsenligt och tämligen flexibelt sätt i förhållande till när nyttan av den immateriella rättigheten erhålls. En enskild konsument betalade tidigare för upphovsrätt när hen köpte en skiva eller en bok men betalar nu i många fall via ett företag som erbjuder strömmande lyssning, som i sin tur betalar upphovsrättsinnehavaren i förhållande till användning. En liknande modell användes redan tidigare för biblioteksersättningen till författare: en viss del av betalningen för upphovsrätten var knuten till inköp av boken och en viss del till utlåning.

Att engångsbetalningar knutna till en vara ersätts av betalningar över tid, en prenumerationsmodell, är en trend som finns på många områden. Mjukvara och andra IT-tjänster säljs i ökande utsträckning genom prenumeration eller andra varianter av *X-as-a-service*. Molnlösningar erbjuder lagrings- och processorkapacitet som en tjänst, istället för att kunden köper egen hårdvara. Uthyrning i olika former är förstås en gammal affärsmodell som innebär att en tjänst säljs, istället för en vara.

Inom offentlig upphandling används ibland funktionsupphandling. Som idealtyp gäller i en funktionsupphandling att det är endast resultatet av upphandlingen som specificeras; de konkurrerande budgivarna är helt fria att välja teknik, material och metod för att nå målet. Funktionsupphandlingens motsats är en traditionell och hårt specificerad upphandling, där budgivaren/leverantören inte har några frihetsgrader utan där köparen bestämt hur den

³⁰ Om ersättningen ska utgå "för evigt" används lämpligen en konstruktion med royalty för ett varumärke eller liknande som kvarstår i säljarens ägo.

³¹ www.prv.se, läst 18 februari 2021.

upphandlade produkten ska se ut och hur den ska produceras. Motsatsparet är kanske tydligast när det gäller vägbyggen.³²

Om en väg upphandlas genom funktionsupphandling bestämmer köparen i det idealtypiska fallet mellan vilka punkter vägen ska dras, vägens bredd, hållbarhet och en del andra kvalitetsaspekter, såsom kurvatur, avrinning, belysning. Kort sagt bestämmer köparen vägens kvalitet i en rad avseenden, men uttalar sig i övrigt inte om *hur* vägen ska byggas. Traditionellt upphandlas vägar istället genom utförarentreprenad. Vid en sådan genomför beställaren projektering, vilket innebär att tekniska konsulter och arkitekter anlitas, som i sin tur specificerar exakt hur vägen ska byggas, med vilket material och så vidare. Potentiella utförare får ta del av projekteringen och lämnar utifrån den anbud.

I princip innebär utförarentreprenad att vägen kommer att se exakt likadan ut, oavsett vem som vinner. Med en funktionsupphandling kan vägen göras på olika sätt; i princip kan olika potentiella utförare tävla om smartare sätt att bygga en väg, som dessutom är anpassade till deras egen tekniska kompetens och deras tillgång till produktionsresurser. Nackdelen med en funktionsupphandling är att vissa kvalitetsaspekter kan visa sig först efter mycket lång tid. Om utföraren slarvar med underarbetet kanske vägen bara håller hälften så länge som den annars skulle gjort, men detta kan vara svårt för beställaren att genomskåda. Utföraren kan hävda att detta är ett smartare sätt att uppnå samma funktionella krav. Anläggningsföretaget sänker sina kostnader och därmed blir en mindre samvetsgrann utförare mer konkurrenskraftig.

En möjlig lösning på dilemmat är att utföraren görs ansvarig för att produkten håller vad den lovar. Om beställaren krävt och utföraren lovat att vägen ska hålla i 40 år, men den behöver byggas om redan efter 20, ska utföraren stå för kostnaden. Emellertid riskerar beställaren att hamna i rättsliga tvister om exakt vilket ansvar utföraren har – och utföraren kan vid den tiden ha gått i konkurs eller på annat sätt försvårat ett sådant ansvarsutkrävande. Bland annat för att minska risken för denna typ av *moral hazard* används ibland ett Offentligt-privat partnerskap, OPS, på engelska kallat Public-Private Partnership, PPP. Genom att utföraren (och dennes konsortium) även står för finansiering minskar möjligheten att smita från ansvar. Nackdelen med OPS-modeller är förstås att privata företag vanligen har högre avkastningskrav än staten, vilket innebär att de finansiella kostnaderna för projektet stiger, möjligen mer än de kostnadsbesparingar som kan göras genom att smartare lösningar väljs.³³

Ytterligare en betalningsmodell som brukar användas är garantier. Säljaren garanterar en viss kvalitet. När den garantin sträcker sig över lång tid används ibland en tredje part för att skapa säkerhet kring att garantin faktiskt kommer att gå att utnyttja; ett exempel på detta är garantier för nybyggda hus.

Sammanfattningsvis finns det många situationer där köparens värde av en produkt är osäker, antingen beroende på osäkerhet om produktens kvalitet eller på osäkerhet om hur mycket köparen kommer att använda produkten, eller både och. Det finns ett antal avtals- och ägarrettsliga lösningar på problemet.

Ägandet till produkten kan kvarstå hos den potentiella säljaren och den potentiella köparen kan disponera produkten genom hyres- eller leasingavtal. En variant är att äganderätten vid

³² Se Eriksson och Hane, 2014 samt Nilsson, 2008.

³³ Nilsson, 2008.

en senare tidpunkt övergår till köparen, exempelvis efter en inledande leasingperiod eller genom en PPP-lösning.

Ägandet av en viss aspekt av produkten, en immateriell rättighet, kan kvarstå hos säljaren, som då kan få royalty som kan vara proportionell mot användningen. Även utan en sådan kvarstående äganderätt kan betalningen vara betingat på utfallet (nyttan) av produkten, till exempel vid företagsöverlåtelser där priset är betingat på framtida resultat eller vid funktionsupphandlingar.

Slutligen kan garantier och bonus användas, som innebär antingen en viss återbetalning till köparen om utfallet är sämre än förväntat, eller omvänt en tilläggsbetalning till säljaren om utfallet överstiger ett visst referensvärde.

7.2 Juridiska förutsättningar för betalningsmodeller för ATMP

Enligt nuvarande regler förväntas ATMP köpas av regioner i ramavtal. Eftersom det i flertalet fall förväntas finnas giltiga patent kan förhandlat förfarande användas. I vissa fall kan en konkurrenssituation ändå föreligga, nämligen om det finns olika ATMP med liknande effekter för samma indikation eller om det finns traditionella läkemedel som kan användas för samma indikation.

Ramavtal sluts för som längst fyra år, enligt LOU, eller längre om det finns särskilda skäl. Ramavtal används när det inte går att exakt förutsäga hur stor förbrukningen kommer att bli. Med grund i ramavtal sker sedan successiva avrop, som för i varje fall ATMP för sällsynta sjukdomar kan förväntas gälla enskilda patienter. Avropen leder till att avropsavtal ingås med leverantören. Avtalslängden för avropsavtalen är inte begränsad till fyra år, men om ramavtalet ingås med en enda leverantör ska avtalsvillkoren för avropsavtalen i allt väsentligt, förutom inköpsvolym, fastställas redan i ramavtalet.

Om ramavtal ingås med flera leverantörer behöver inte alla villkor för avropsavtalen fastställas. I en förnyad konkurrensutsättning, där enbart ramavtalsparterna får delta, kan exempelvis priset för ett visst avrop bestämmas.

Den ekonomiska risken för regionerna kan begränsas på olika sätt. Ramavtal kan ingås för korta perioder, exempelvis ett eller två år, istället för fyra år eller längre. Då hinner färre patienter behandlas och priset kan sänkas om det medicinska utfallet visar sig vara sämre än förväntat. Ramavtalen kan även begränsas av tak för hur många avrop som kan ske.

Om det finns andra läkemedel – ATMP eller traditionella – som kan användas för samma indikation kan priset pressas, antingen i den initiala upphandlingen eller vid förnyade konkurrensutsättningar när det blir aktuellt med avrop.

En viktig fråga, som vi återkommer till nedan, är om betalningen ur ekonomisk synpunkt bör utgå i form av ett engångsbelopp eller fortlöpande, i takt med att hälsonytta realiserar. Vi saknar den juridiska expertisen för att kunna avgöra om det är möjligt att betala delar av ersättningen för ATMP lång tid i efterhand. Vår bedömning är dock att detta i varje fall inte strider mot upphandlingsreglerna och att det antagligen går att hitta en avtalskonstruktion som skulle tillåta att delar av betalningen faller ut med en lång fördröjning.

Av praktiska skäl tror vi att det är mer ändamålsenligt att avropsavtalet innehåller bestämmelser om tilläggsbetalningar som är kopplade till något utfallsmått, än bestämmelser om garanti eller andra bestämmelser om partiell återbetalning.

Vi har tagit del av bedömningar som pekar mot att det är svårt att på ett objektivt sätt fastställa det medicinska utfallet för enskilda patienter. Dock kan en fortlöpande betalning betingas på om patienten lever eller inte. Vår bedömning är att det därtill skulle vara möjligt att utvärdera det långsiktiga medicinska utfallet av ATMP på populationsnivå. Exempelvis skulle TLV kunna få i uppdrag att vart femte år revidera den hälsoekonomiska bedömningen av ett ATMP. En sådan revidering kan ta hänsyn till såväl nya studier av läkemedlets effekt som till bättre alternativa behandlingar och prissänkningar för andra läkemedel på grund av patentutgångar eller ökad terapeutisk konkurrens. Det är av särskild vikt att TLV reviderar bedömningen av ATMP i de fall där ICER ligger nära tröskelvärdet, med hänsyn tagna till rabatter i den regionala upphandlingen.

Att ett företags prissättning på detta sätt begränsas av en myndighet är inte unikt. I praktiken är ju detta vad TLV gör för traditionella läkemedel. Även på andra områden, som exempelvis marknaderna för tele och el, begränsas prissättningen av myndigheter.

Det bör dock övervägas om den här typen av effektbedömningar bör göras på EU-nivå, vilket ju i stor utsträckning är fallet vad gäller telereglering.

7.3 Betalningsmodeller vid asymmetrisk information. Vad kan vi lära av teorin för *mechanism design*?

Vi har ovan analyserat hur den förväntade hälsovinsten ska beräknas vid osäkerhet. Ett problem som vi indirekt berört har att göra med att den ena parten (läkemedelsföretaget) har mer information än den andra (myndigheterna). Osäkerheten kring effekten av ATMP kompliceras alltså av att det råder *informationsasymmetri*. Läkemedelsföretaget har en bättre kunskap om ett specifikt läkemedel än det offentliga. När det råder osäkerhet om värdet för en produkt och när informationen är asymmetriskt fördelad finns olika avtalsmodeller som brukar användas. Typiskt sett går dessa modeller ut på att ansvaret för osäkerheten läggs på den part som har bäst informationsläge – vilket normalt är säljaren. Enligt redogörelsen i avsnitt 7.1 kan det ske genom att säljaren erbjuder garanti eller genom att betalningen erläggs successivt, i takt med att nyttan uppstår. Det senare är en vanlig modell för immateriella rättigheter, enligt diskussion ovan.

Skälet till att lägga ansvaret på den bättre informerade parten är att denna annars kan exploatera den sämre informerade parten, exempelvis genom att avslöja information som är gynnsam för parten själv, men inte avslöja information som inte är gynnsam. Den part som har sämst information kan hantera situationen genom att erbjuda motparten en meny av avtal (eller priser), där menyn är utformad så att motparten har incitament att avslöja sin information.

Ett exempel på detta är försäkringar, där hög självrisk ger lågt pris. En försäkringstagare som vet med sig att hen har hög risk kommer att avslöja detta genom att välja en försäkring med låg självrisk och får då betala ett högre pris. Ett annat exempel är en arbetsgivare som erbjuder antingen en måttligt hög fast lön eller en låg fast lön i kombination med en rörlig ersättning som totalt sett kan ge en hög lön för den arbetstagare som har god förmåga och hög kapacitet. Ytterligare ett exempel är ett företag som erbjuder ett högt pris för "all-you-can-eat", för kunder som vill (och kommer att) konsumera mycket av företagets tjänster, liksom ett lågt startpris i kombination med ett ganska högt rörligt pris, för kunder som vill

(och kommer att) konsumera litet. Metoden tillämpas exempelvis av telebolag, nöjesparker och restauranger.

Nedan diskuterar vi om och hur principen – som också kallas *mechanism design* – skulle kunna tillämpas för ATMP-behandlingar. I korthet innebär principen i detta fall att den initiala ersättningen för ATMP bara eller i huvudsak baseras på effekter som tillverkaren kunnat påvisa med mycket hög säkerhet – men att det i ersättningsmenyn finns möjlighet att få en totalt sett högre ersättning om och när det visat sig att effekten var stor.

Detta sätt att hantera informationsasymmetrin ger den som har (eller kan skaffa) mer information incitament att korrekt informera (och skaffa ytterligare information). Hur en principal (huvudman) ska utforma spelregler för agenter som har bättre information som gör det optimalt för agenterna att avslöja sin information är ett exempel på ett *mechanism design* problem. Ett exempel som vi nämnt ovan är ett försäkringsbolag som kan erbjuda låg självrisk till ett högt pris och hög självrisk till ett lågt pris. Om prissättningen är korrekt utformad i förhållande till olika agents försäkringsrisk kommer dessa att välja "rätt" självrisk och därigenom avslöja om deras risk är hög eller låg.

Vi tänker oss att tillverkaren av ett ATMP har bättre information om läkemedlets effekter än myndigheterna. Tillverkaren kan strategiskt välja att avslöja eller inte avslöja information, men det kan inte ljuga. Vårt motiv för detta är att vi tror att ett läkemedelsbolag som fabricerar bevis riskerar mycket allvarliga sanktioner och dessutom skulle förlora trovärdighet mot sina kunder.

För att ge tillverkaren starka incitament att för det första skaffa tillförlitlig information om ett nytt ATMP:s långsiktiga effekter och för det andra att avslöja all relevant information som framkommit, även negativ sådan, bör det offentliga erbjuda en (engångs)ersättning som är så låg att sannolikheten för att utfallet ska bli ännu sämre än vad den låga ersättningen motiverar är mycket liten.

Med en sådan ersättningsmodell kommer det att bli attraktivt för läkemedelsföretaget att acceptera en kontinuerlig betingad betalning. Det är svårt att mer exakt slå fast hur en låg engångsersättning ska beräknas, men det skulle exempelvis kunna vara den ersättning som är motiverad om utfallet blir motsvarande den undre gränsen i ett 95-procentigt konfidensintervall för skattad effekt.

Med en sådan ersättningsmodell bör risken för negativa överraskningar vara begränsad, samtidigt som tillverkaren har starka incitament att skaffa mer information. I takt med att konfidensintervallet krymper, det vill säga i takt med att informationen blir mer precis, stiger ersättningen. Det kommer också att bli attraktivt för en tillverkare som tror på sin egen produkt att välja en ersättningsmodell som är betingad på utfall.

Vi har inget särskilt argument för att just gränsen för ett 95-procentigt konfidensintervall ska väljas; detta är ett relativt godtyckligt val.³⁴ För en vidare diskussion om betalningar betingade på information, se Appendix E.

³⁴ Staten har valt en modell med liknande egenskaper för reavinstbeskattning av aktier. Reavinsten beskattas med 30 procent av vinsten, där vinsten antingen beräknas som faktisk vinst eller som 80 procent av försäljningssumman (schablonmetoden). Den aktieägare som har glömt inköpspris kan använda schablonmetoden men det kan också den aktieägare som har ett inköpspris som understiger 20 procent av försäljningssumman. Den aktieägare som har ett högre inköpspris än så får ett starkt incitament att hålla reda på faktiskt inköpspris.

Sammanfattningsvis är vår bedömning att prissättningen bör ge företagen starka incitament att avslöja så mycket information som möjligt. Detta talar för att TLV bör fastställa konservativa bedömningar av sannolikheter som är förenliga med den information som finns. Detta ger även företagen incitament att, i de allra flesta fall, avslöja all information de har, men det ger också företagen incitament att skaffa information. Vid bedömningen av dessa osäkerheter bör också förväntad utveckling av jämförelsealternativet beaktas.

En konservativ modell för en engångsersättning för ATMP ger vidare tillverkarna incitament att föredra en fortlöpande ersättning som är betingad på utfall. I själva verket kan ett huvudsyfte med den konservativa beräkningsmodellen för engångsbetalning vara att ge tillverkarna incitament att acceptera en betingad fortlöpande betalning. Vi fortsätter att diskutera den frågan i följande avsnitt.

7.4 Behandlingsneutralitet och information

Som vi visade i kapitel 4, innebär värdering av ATMP med förskottsbetalning att framtida osäkerheter långt fram i tiden måste värderas. Inlåningen innebär att osäkerheten om de långsiktiga hälsoeffekterna av ATMP måste bedömas. Om ATMP innebär fullgod hälsa, vad är sannolikheten att den på lång sikt inte är det? Eftersom det, som vi understrukt ovan, inte är möjligt att få bättre än full hälsa innebär det en övervärdering. Även en liten sannolikhet för att hälsoeffekten upphör påverkar i hög utsträckning en värdering om den görs över decennier av liv. För en kontinuerlig behandling är detta problem vanligtvis mycket mindre allvarligt, eftersom sjukvården då upphör med betalning och behandling.

För behandlingar där förväntad överlevnad utan ATMP är hög är kanske den främsta osäkerheten i värderingen av det kontrafaktiska alternativet. Hur skulle hälsa och kostnader se ut för patienter med bästa alternativa behandling kommande decennier?

Ett extremfall som illustrerar betydelsen av analys av det kontrafaktiska alternativet är en ATMP som framför allt syftar till att sänka vårdkostnader över lång sikt. Då är det ganska uppenbart att en god prognos för hur de faktiska kostnaderna förväntas utvecklas är nödvändig för att bedöma om insatsen är motiverad. I dessa fall är det av vikt att företaget på ett övertygande sätt visar att det inte förväntas kostnadsänkningar för befintlig behandling av motsvarande storlek under den relevanta tidshorisonten. I och med att prognoser över ekonomisk utveckling över mer än ett fåtal år är mycket osäkra i allmänhet, bör rent kostnadsänkande ATMP bedömas över en relativt kort tidshorisont.

7.5 Betalningar betingade på utfall

Det finns alltså mycket som talar för utfallsbaserad ersättning. Som diskuterats i kapitel 4 innebär full ersättning till företaget vid botande betalning en förskottsbetalning för en osäker patientvälfärd. För att inte övervärdera ATMP måste hänsyn tas till osäkerhet om de långsiktiga effekterna av läkemedlet och den långsiktiga utvecklingen av värdet av den kontrafaktiska behandlingen. Genom att betinga ersättning på utfall kan samhället göra ersättningen mer behandlingsneutral.

Givet att företagen är mycket mer informerade om risker torde det vara önskvärt att de tar på sig så stor andel av risken med ATMP som möjligt, även med hänsyn till att staten generellt sett, genom sin större ekonomiska styrka, egentligen har bättre kapacitet att bära

ekonomiska risker. Betalningen bör alltså på något sätt betingas på utfallet, antingen på individ- eller populationsnivå. Det kan ske antingen genom att betalningen betingas på det hälsoutfall som man verkligen är intresserad av eller genom att man betingar på någon variabel som är korrelerad med utfallet, en så kallad *proxy*.

Den värdering staten gör baseras i stor utsträckning på det underlag som företagen lägger fram. Värderingen görs också med begränsade resurser över en kort tidsperiod. Mycket av underlaget är dessutom sekretessbelagt, vilket försvårar en djupare analys och kunskapsuppbyggnad om de långsiktiga effekterna av ATMP. Företaget som utvecklat en behandling och analytiker på finansiella marknader har rimligtvis en bättre kunskap om osäkerheter mm. Det talar för att det offentliga bör försöka utforma kontrakt där ersättning kopplas till utfall.

För att se hur analysen påverkas, antag att företaget vet med säkerhet om behandlingen är värd att genomföra, men det offentliga inte vet. Med förskottsbetalning väljer vårdgivaren att behandla om det förväntade nuvärdet (dvs NB) givet osäkerheten är positiv. Det innebär att det finns en positiv sannolikhet att man behandlar även när behandlingen inte bör ske. Om företaget ersätts enbart vid lyckad behandling, kommer det att välja att introducera behandlingen enbart om den är värd att ge.

En annan fördel med betingade kontrakt är att det minskar osäkerheten i värdet av behandlingen med ett sådant kontrakt. Även om företaget och det offentliga har olika uppfattning om hur stor andel som botas, behöver man inte bli överens om en gemensam uppskattning för att bestämma värdet av behandling *givet* att behandlingen lyckas.

Ett viktigt argument mot betalningar över tid är att det kan vara svårt att påvisa medicinskt utfall. Det kan behövas dyra undersökningar som är krävande för patienterna och patienterna kan vägra att medverka. Vi menar emellertid att det bör utredas om det kan vara möjligt med en betalningsmodell som bygger på en årlig betalning så länge patienten är vid liv, som administreras (och möjligen betalas; se nästa avsnitt) av staten. Det årliga belopp som betalas fastställs i samband med upphandlingen av ATMP för en period av exempelvis tre till fem år. Därefter justeras den hälsoekonomiska beräkningen av TLV enligt de principer som används för läkemedelsförmånen.

För enkelhets skull tänker vi oss här att osäkerheten består i livslängd. En möjlighet är att det offentliga ersätter företaget vid behandlingen baserat på förväntad livslängd, med korrigerande betalning om utfallet avviker från det förväntade. Ett alternativ är att ersättning sker löpande, baserat på utfall. I det första fallet har vårdgivaren en eventuell fordran, i det andra företaget.

En fordran på staten är mer säker än en fordran på ett företag, eftersom det är större sannolikhet att företaget ställer in betalningarna. Därför är det troligt att staten kan skjuta upp betalningar mot löfte om betalning i framtiden. Om företaget skulle gå i konkurs skulle denna fordran övertas av annan ägare. Om däremot staten har en fordran på företaget är risken stor att denna går förlorad, eller bara ersätts delvis, vid en konkurs. Problemet förvärras av möjligheten för den ursprungliga tillverkaren att sälja en substans till ett företag med mindre ekonomiska muskler, så fort det börjar framstå som sannolikt att läkemedlet inte kommer att motsvara förväntningarna.

Ett sätt att hantera osäkerheten i utvecklingen av det kontrafaktiska alternativet är att låta ersättningen bero på hälsovinsten vid aktuell tidpunkt. Det skulle kunna utformas på olika sätt, men ett sätt är koppla den till förnyade värderingar.

Hälsovinsten beräknas av TLV inkrementellt i förhållande till bästa alternativa behandling på sedvanligt sätt; den således beräknade hälsovinsten jämförs med den beräkning som låg till grund för det ursprungliga ramavtalet. Man kan se det så att ersättningen fastställs i en upphandling, men indexeras till bästa aktuella kunskap om hälsoekonomisk vinst, med exempelvis tre- eller femåriga revisionsperioder.

För att ta ett konkret exempel. Anta att hälsovinsten beräknas till 0,5 QALY per år, så länge patienten är i livet. I upphandlingen av ramavtal fastställs priset per år till 300 000 kronor. Vid revisionen efter fem år pekar bästa tillgängliga evidens på att hälsovinsten, med beaktande av det vid den tiden bästa alternativet, bara är 0,3 QALY per år. Ersättningen justeras då till $(0,3/0,5) \times 300\,000 = 180\,000$ kronor per år.

7.6 Kostnader och svårigheter förknippade med utfallsbaserad prissättning

I praktiken finns det problem med utfallsbaserad ersättning. I vissa fall har det visat sig vara administrativt mycket kostsamt att genomföra, exempelvis behandlingen av MS i England. Ett annat exempel, i Nederländerna, stötte på problemet att det var svårt att sluta subventionera ett läkemedel som en gång godkännts, även om det visade sig fungera sämre än förväntat och därför inte längre var kostnadseffektivt. (OECD, 2017.) Liknande problem har upplevts i Sverige för läkemedlet Duodopa; se nedan.

Vissa utfallsbaserade mått är enkla att använda. Registerdata hos myndigheterna, exempelvis om patienten är vid liv är enkelt att använda i automatiserade rutiner. Det kan fånga en osäkerhet om förväntad livslängd. Om det finns osäkerhet om livskvalitet efter behandlingen krävs dock andra mått. Annan registerdata som t.ex. om patienten är arbetsför eller inte fångar bara delvis hälsostatus.

Bättre indikatorer på hälsostatus går att ta fram, men detta är kostsamt eller opålitligt eller både och. Patienterna har kanske inte incitament att medverka eller avslöja sin hälsostatus.

Läkarna har ofta/ibland incitament att också förskriva traditionella läkemedel, vilket försvårar bedömningen av värdet av engångsbehandlingen. Ett alternativ är att använda populationsdata. Vi menar att TLV kan utveckla rutiner för att göra samma typ av bedömning exempelvis vart tredje eller femte år, som myndigheten gör när läkemedlet först börjar säljas.

Hanna med flera (2018) redovisar en litteraturstudie av de förslag som presenterats i 48 tidskriftsartiklar som behandlar finansiering av ATMP. Författarna delar in förslagen i tre huvudkategorier: finansiella överenskommelser, hälsoutfallsbaserade överenskommelser respektive "hälsomynt". Finansiella överenskommelser, den vanligast förekommande kategorin, handlar exempelvis om (hemliga) rabatter, volymbegränsningar, volymberoende priser, kostnadsbaserad (till skillnad från exempelvis värdebaserad) prissättning, "avbetalningsköp" (till exempel annuitetsbetalningar) eller innovationspriser (*innovation prizes*). Överenskommelser baserade på hälsoutfall kan vara betingade antingen på individuella utfall eller populationsutfall. "Hälsomynt", slutligen, tar sikte på att fördela kostnaden mellan olika försäkringsbolag och/eller staten, till exempel om en behandlad

patient flyttar eller uppnår en ålder då ansvaret för sjukvårdskostnader övergår från ett privat försäkringsbolag till staten (exempelvis via Medicare i USA).

Författarnas slutsats är att annuitetsbetalningar och motsvarande troligen inte är rätt väg att gå, utan att snarare att en "ATMP-specific fund" bör användas för att jämna ut betalningar över åren, med tydligt definierade kriterier för vilka läkemedel som får finansieras med hjälp av denna. Författarna bedömer vidare att det är nödvändigt med mekanismer för att begränsa kostnaderna och att då en "CED [Coverage with Evidence Development] with escrow agreements", det vill säga betalningar betingade på populationsdata.

För att hantera problemet att det är svårt att sluta subventionera ett läkemedel som senare visar sig inte vara kostnadseffektivt föreslår de att CED används i kombination med ett depositions- eller bankgarantiavtal. Detta, menar de, är ett bra sätt att både stimulera till utveckling av nya läkemedel och göra läkemedelskostnaderna ekonomiskt hållbara.³⁵ Genom depositionsavtalet kan läkemedelsbolagen bindas till att återbetala en del av ersättningen om och när det visar sig att effekten är sämre än förväntat. (Se vidare nedan.)

Hanna med flera refererar fyra artiklar som i sin tur behandlar olika varianter av CED: Fellows och Hollis, 2013; Garrison med flera, 2015; Stafinski med flera, 2010; och Towse och Garrison, 2010.

Persson med flera (2012) föreslår en kombinerad finansiering, där staten betalar en värdebaserad priskomponent medan regionerna betalar en kostnadsbaserad ersättning. Hanna med flera menar dock att kostnadsbaserad ersättning möjligen är den sämsta av alla modeller, då den enligt dem skulle avskräcka från investeringar. Vi har svårt att se poängen med förslaget, annat än att det (ur läkemedelsbolagens synvinkel) skulle göra det möjligt med högre ersättningar.

Dabbous med flera (2020) diskuterar de praktiska erfarenheterna av olika typer av *Managed Entry Agreements* (MEA) som prövats i Europa. CED har av flera bedömare ansetts vara principiellt ändamålsenlig, men har visat sig vara praktiskt svårt att tillämpa i den form som varit vanlig: att offentlig finansiering upphör för läkemedel som efter ytterligare studier inte bedöms vara kostnadseffektiva. När detta har prövats – exempelvis i Sverige för läkemedlet Duodopa – uppstod ett stort tryck på myndigheterna att ompröva detta beslut.

För att hantera detta problem har Frankrike prövat en CED där betalningen deponeras i en bank i väntan på resultat (*escrow agreement*) för läkemedlet Risperdal Consta. Svårigheterna att använda CED har fått de flesta europeiska länder att istället använda MEA som tar sikte på finansiella överenskommelser (se referatet av Hanna med flera ovan).

³⁵ Författarna skriver sammanfattningsvis att "CED appears to be the right tool to address uncertainty, it helps to address the growing concerns over the quality of clinical evidence submitted at time of decision making and make informed decisions on value of the new therapy but it is unlikely to have an impact on the affordability and the healthcare spending trends [...]. It may fail in some cases to capture the data needed to reduce the uncertainty associated with decision-making; thus, the agreement may be terminated. In the Netherlands, the experience with CED for orphan drugs was not favorable; some drugs were not proven cost-effective, but it was impossible to stop coverage due to public pressure [...]. Furthermore, it may have operational challenges in practice such as insufficient budget to fund data collection or inadequate data collection systems [...] CED should be preferred [över betalningar betingade på individuella medicinska utfall] whenever possible. A CED scheme may be complemented with an escrow agreement. The escrow agreement places the sales revenue in an independent bank account; at the end of the study, the collected money is released to the company if the results are positive and to the health insurance if the results are negative.

Författarna bedömer, trots detta, att "To inform value-based decision making and allow early access to innovative medicines, CEDs should be revisited."

8 Ansvarsfördelningen mellan stat och regioner

- *Slumpvariationer i antalet patienter som förskrivs med dyra ATMP förväntas leda till relativt små kostnadsfluktuationer, i förhållande till regionernas totala budget. I förhållande till deras läkemedelsbudget kan det dock bli relativt stora fluktuationer.*
- *En större utmaning är dock tillfälliga men kraftiga kostnadsstegringar när nya ATMP lanseras för stora patientgrupper. Här bör dock staten kunna ge tillfälliga stöd.*
- *För det fall fortlöpande betalningar betingade på utfall på populationsnivå används (CED, coverage with evidence development) talar administrationskostnader för en statlig lösning.*
- *Om ATMP framför allt utgör substitut för traditionella sjukvårdsläkemedel talar detta för att kostnadsansvaret ska ligga på regionerna; om ATMP framför allt utgör substitut för traditionella förskrivningsläkemedel talar detta för att kostnaderna ska bäras av staten.*
- *Den samverkande effekten av värdebaserad prissättning, budgeteringsprinciper och regelverken i övrigt gör att regionerna har ett större kostnadsmotstånd, varför de sannolikt kan få lägre priser än vad som vore möjligt med en statlig lösning.*
- *Staten har större förmåga att bära stora kostnadsvariationer och att betala i förskott för framtida hälsonytta; om betalningsmodellen för ATMP bygger på engångsbetalningar talar detta för att staten ska ta kostnadsansvaret.*
- *Det finns således argument såväl för statligt som regionalt kostnadsansvar; vilket som är att föredra beror på vilken tyngd som ges åt de olika argumenten ovan.*

Stat och regioner samarbetar vid bedömning och upphandling av ATMP. En förväntad ökning av ATMP-behandlingar innebär att det finns skäl att fundera över om nuvarande arbetsfördelning är ändamålsenlig. Det finns mycket som talar för behovet av utökad samarbete, alternativt en utökad statlig roll.

8.1 Riskfördelning

Det förefaller sannolikt att samhället kommer att få hantera relativt stora variationer i kostnader för ATMP-behandlingar. Dessa risker kan förmodas vara relativt stora eftersom behandlingarna är dyra och eftersom sjukdomarna är ovanliga. Variationen är även större om ersättningen görs i form av engångsbelopp. Detta kan förmodas leda till stora svängningar i ATMP-betalningarna från år till år, särskilt i små regioner. För en region innebär detta en budgetär risk, i den mån variationen är kostsam. Som andel av en regions totala budget torde sådana variationer ofta vara relativt små, men som andel av läkemedelsbudgeten kan de vara betydande.

Den budgetära kostnaden för en region av en ATMP-behandling beror på ett antal faktorer utöver betalningsform över tiden. Behandling av Hepatit C innebar exempelvis stora utmaningar för regionerna, i och med att det fanns relativt många som skulle behandlas. Det innebar en engångsökning av kostnaden i och med att stocken av konstaterade fall är större än flödet av nya fall.

Denna engångskostnad som uppstår av att börja behandla konstaterade fall är dock mindre för de sjukdomar där hälsovinsten är som störst för ATMP. I dessa fall innebär ATMP ett långt liv snarare än tidig död. Allvarligheten innebär då också att det finns färre konstaterade fall att behandla vid införandet av ATMP, vinsten och kostnaden kommer att vara spritt över tiden.

På riksnivån blir dessa variationer av mindre betydelse på grund av de stora talens lag. Detta talar för att staten och inte regionerna bör ta risken. Detta kan antingen ske genom att staten tar på sig ansvaret för ATMP-läkemedel eller genom att man tillhandahåller någon form av försäkringsmekanism till regionerna.

Det kan finnas gränsdragningsproblem för vilken vård som stat och region bör svara för. En fördel med ett frivilligt samarbete mellan regionerna som NT-rådet, är att samarbete kunde uppstå när det ger stora fördelar utan att behöva formalisera omfattningen av samarbetet.

8.2 Efterfrågan (patientbehov)

Dyra ATMP kan innebära en inte obetydlig aggregerad risk/osäkerhet från kostnadssidan om i) sjukdomen som ATMP används för är geografiskt koncentrerad (Skellefteåsjukan), ii) det förväntade antalet fall per år i en region är litet men ett mått på variationen i antal fall gånger priset ändå är så stor att det i förhållande till regionens budget är ett betydande belopp eller iii) det uppstår variationer över tid i takt med att olika ATMP lanseras för olika patientgrupper.

Angående ii) så handlar det om risk, inte en osäkerhet, och just därför är det relativt enkelt att göra beräkningar. Två räkneexempel kan illustrera principen.

En sjukdom som inträffar i förväntan en gång per år i en befolkning på 100 000 individer observeras med ungefärliga sannolikheterna 37, 37, 28 respektive 6 procent noll, en, två respektive tre gånger under ett år. Om ATMP kostar 20 Mkr kan vi approximera detta med att vi med ca 1/3 sannolikhet får en kostnad på 20 miljoner kronor, med en lika stor sannolikhet en kostnad på 40 miljoner kronor och med en lika stor sannolikhet en kostnad på 60 miljoner kronor – eller möjligen mer. Regionernas budget är i storleksordningen 10 procent av BNP eller 5 miljarder kr per 100 000 invånare så en variation på plus/minus 20 miljoner kronor motsvarar mindre än en halv procent av den årliga budgeten. Dessutom finns ju inga regioner som är så små och vid större populationer kommer fördelningen att vara mer koncentrerad kring förväntat värde.

Om sjukdomen istället inträffar en gång på 10 000 är, det vill säga i förväntan 10 gånger i en lika stor befolkning som i exemplet ovan, kan standardavvikelsen för utfallet av antal sjukdomsfall per år beräknas bli cirka 3 eller 60 Mkr.³⁶ I förhållande till en budget på 5 miljarder kronor är detta litet mer än 1 procent. Om läkemedel exempelvis står för 6 procent av regionernas totala kostnader innebär dock detta en femtedel av de årliga läkemedelskostnaderna.

³⁶ Om sjukdomen istället inträffar en gång på 10 000 är, det vill säga i förväntan 10 gånger i en lika stor befolkning som i exemplet ovan, kan standardavvikelsen med formeln

$$\sqrt{N \cdot P \cdot (1 - P)}$$

där N är befolkningen och P är sannolikheten för sjukdomsfall. Med antaganden enligt ovan blir standardavvikelsen cirka 3 eller 60 Mkr.

Om regionen istället har en miljon invånare ökar standardavvikelsen till 10 vilket med samma antaganden som ovan motsvarar 6 procent av de årliga läkemedelskostnaderna, som i sin tur utgör 6 procent av regionens totala kostnader.

Angående punkten iii) ovan, risken att en ny stor patientgrupp kan behandlas med ATMP, kan detta potentiellt bli ett stort budgetproblem om det nya läkemedlet är dyrt. En observation är dock att förutsättningen för att det ska finnas en stor population potentiella användare av en ATMP är att det antingen gäller en sjukdom som inte medför en stor risk för dödsfall inom kort tid, eller gäller en sjukdom för vilken det redan finns alternativa behandlingar. Två exempel som kan nämnas, som medfört stora kostnader under en period, är vaccin för covid-19 respektive HPV. Regeringen har avsatt 4,5 miljarder kronor för inköp av covid-19-vaccin.³⁷ Kostnaden förväntades i december stanna vid 1,2 miljarder kronor, men därtill kommer kostnader för regionerna att administrera vaccin, vilket med en ersättning till regionerna på 550 kronor per person blir en hög kostnad.³⁸

En lösning på problemet är att patienter sätts på kö för att få ATMP, en annan att staten går in med stöd, en tredje att regionerna helt enkelt får ta en sådan kostnad. Ytterligare en lösning är att betalningen sker som en "ström" över tiden, istället för genom en engångsbetalning.

8.3 Engångskostnader för ATMP

När ATMP introduceras kan kostnaderna stiga för det offentliga den första tiden, i och med att det kan finnas ett antal patienter att behandla direkt. När väl dessa har behandlats sjunker behandlingskostnaderna i och med att nyttillkomna fall är mycket färre. Introduktionen av behandling av Hepatit C är ett exempel på detta, där regionerna fick extra medel för att klara av de extraordinära kostnaderna.

Det många sjukdomar där behandlingar skulle innebära mycket höga kostnader för samhället initialt. Behandling av MS, Huntingtons sjukdom eller diabetes typ 1 med ATMP skulle kunna innebära stora finansieringsproblem med nuvarande system. I och med att behandlingar av sådana sjukdomar inte uppstår så ofta, kan finansieringsfrågan kanske lösas från fall till fall. Därför fokuserar vi här på finansiering av ATMP för mycket ovanliga sjukdomar.

I och med att ATMP kan innebära stora hälsovinster kan också priset på dessa bli högt, vilket skulle kunna leda till mycket höga utgifter vid dess introduktion. Problemet med att finansiera introduktionen av ATMP behandlingar behöver dock inte vara fullt så stort, av det enkla skälet att om en ATMP behandling har ett mycket högt värde är det ofta på grund av att sjukdomen utan behandling är dödlig. Det innebär samtidigt att det inte finns så många patienter att behandla initialt.³⁹

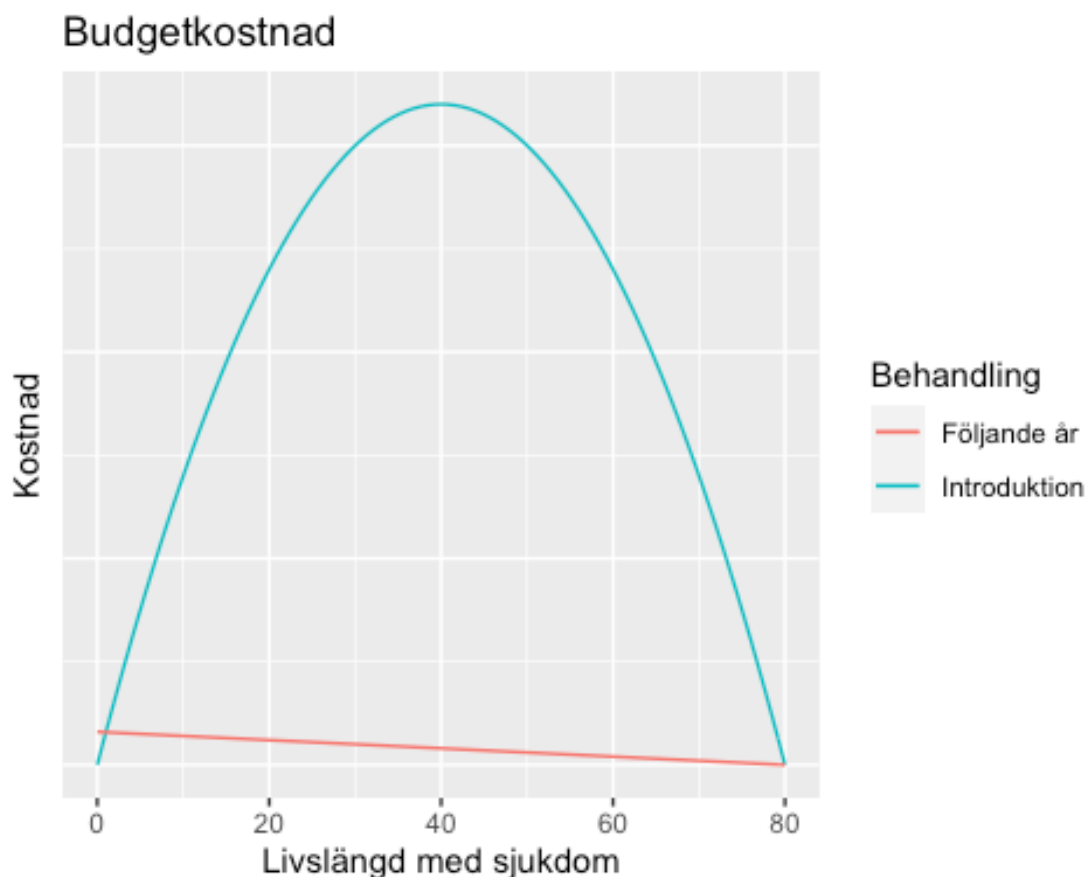
³⁷ Folkhälsomyndigheten, 8 december 2020.

³⁸ DN, 21 november samt 2 december 2020. Regionerna får en ersättning från staten på 550 kronor per person som fått två doser.

³⁹ Vi gör en enkel kalkyl för att grovt belysa problemet med engångskostnader. Antag att det föds N personer varje år. Vi antar att sjukdomen är genetisk, individer föds med den med en liten sannolikhet P_g . Alla individer som inte har sjukdomen lever T år, men om man har den lever man utan behandling bara G år. Antalet patienter att behandla vid liv vid en tidpunkt är då

$$N_g = P_g N G$$

Givet sannolikheten att individer bär på anlaget, är den totala initiala kostnaden som störst för genetiska sjukdomar som drabbar först i medelåldern, som exempelvis Huntingtons sjukdom. Utifrån incidens och förväntad levnadstid kan man beräkna förväntade kostnader för olika genetiska sjukdomar. I figuren nedan illustreras initiala kostnader som funktion av ålder. Kurvans höjd beror på tröskelvärde för ICER och sjukdomens incidens. Oavsett dessa är kostnaden som högst för sjukdomar med ca 40 års livslängd utan behandling.



Man kan notera att värdet av behandling för individer som inte behandlas initialt samtidigt blir högre, eftersom de kommer att vara yngre än de som initialt behandlas. I princip skulle en omvärdering av prissättningen kunna leda till ett högre pris.

8.4 Lånemöjligheter

Staten är även bättre på att hantera osäkerheter över tiden, om det uppstår många fall och höga kostnader vissa år. Detta på grund av att man har större möjligheter att låna och att man har fler skatte-instrument.

Värdet av en ATMP behandling som leder till fullgod hälsa är $T - G$ ökade levnadsår $V_g = T - G$. Om läkemedelsföretaget sätter priset så att överskottet ICER villkoret $p_g = IV_g$ är uppfyllt har vi en kostnad per innevånare för behandling

$$N_g p_g / N = N_g IV_g / N = IP_g (T - G)$$

Den totala behandlingskosten vid introduktionen blir G gånger större än senare år, eftersom individer upp till G år är vid liv när behandlingen introduceras.

Staten kan hantera den puckel som torde uppstå vid lansering av nya mediciner då det finns en stor grupp personer med sjukdomen.

Detta talar för att staten och inte regionerna tar på sig att fördela kostnaderna över tiden. Detta kan antingen ske genom att staten tar på sig ansvaret för ATMP-läkemedel eller genom att man tillhandahåller någon form av lånemekanism till regionerna.

8.5 Forskningsincitament

En orsak till värdebaserad prissättning är att ersättningen ska bidra till företagets incitament till att forska fram nya läkemedel. Sådan forskning är en kollektiv nyttighet. Det finns en större risk för att enskilda regioner bortser från detta värde än att staten gör det. Detta talar för att åtminstone den värdebaserade delen av läkemedelspriser (alltså den del som överstiger produktionskostnaderna) bör betalas av staten snarare än av enskilda regioner.

8.6 Administrativ kompetens

I tidigare avsnitt diskuterades möjligheten att införa betalningsmodeller som är utfallsbaserade eller där kostnader sprids över tiden. Dessa betalningsmodeller kan bli komplicerade att hantera administrativt. Utfallsbaserade modeller kräver en utformning utifrån de specifika omständigheter som finns för ett ATMP, exempelvis relevant statistik. Många olika läkemedel, där olika betalningsmodeller används innebär en stor administrativ börda även för en större region. För en liten region kan det vara osäkert om ett ATMP ens kommer att användas under närmaste år.

Dessa administrativa kostnader uppvisar skalfördelar. Det är bättre om dialogen med läkemedelsföretaget och utformningen av kontraktsformer sker på central nivå.

ATMP innebär ofta att behandlingarna kommer att genomföras på några få platser i landet. Det talar för att uppföljning och kontraktering också bör ske centralt.

8.7 Substitut

Ett argument för att regionerna ska ha ansvaret för ATMP är att regionerna kan få incitament att välja bort billiga behandlingar som man får stå för själv och istället remittera patienter till statlig ATMP-vård.

Det motsatta problemet kan dock också vara relevant. Om alternativet till ATMP är egenbehandling med läkemedel inom förmånen, kan regionen ha incitament att inte välja ATMP om det finansieras regionalt.

Ett annat argument kan vara att det blir svårare att få till terapeutisk pris-konkurrens när det är två olika köpare av ATMP och det konkurrerande läkemedlet.

9 Referenser

- Claxton, Karl, Mike Paulden, Hugh Gravelle, Werner Brouwer och Anthony J. Culyer, (2011), "Discounting And Decision Making In The Economic Evaluation Of Health-Care Technologies", *Journal of Health Economics*, 20: 2–15
- Dabbous, M., Chachoua, L., Caban, A., & Toumi, M. (2020). Managed entry agreements: policy analysis from the European perspective. *Value in Health*, 23(4), 425-433.
- Dixit, Avinash K. och Robert S. Pindyck. 1994. *Investment Under Uncertainty*. Princeton University Press.
- Drummond, M., M. Sculpher, K. Claxton, G. Stoddard, G. Torrance. 2015. *Methods for the economic evaluation of health care programmes*. Oxford University Press.
- Eriksson, Per Erik och John Hane, 2014. *Entreprenadupphandlingar. Hur kan byggherrar främja effektivitet och innovation genom lämpliga upphandlingsstrategier?* Uppdragsforskningsrapport 2014:4, Konkurrensverket.
- Garber, Alan M., och Charles E. Phelps, 1997, Economic foundations of cost-effectiveness analysis. *Journal of Health Economics* 16: 1-31
- Gollier, Christian. 2013. *Pricing the Planet's Future: The Economics of Discounting in an Uncertain World*. Princeton University Press.
- Hanna, E., Toumi, M., Dussart, C., Borissov, B., Dabbous, O., Badora, K., & Auquier, P. (2018). Funding breakthrough therapies: A systematic review and recommendation. *Health Policy*, 122(3), 217-229.
- Kyle, Margaret K. "Are important innovations rewarded? Evidence from pharmaceutical markets." *Review of Industrial Organization* 53.1 (2018): 211-234.
- Lakdawalla, Darius N. "Economics of the pharmaceutical industry." *Journal of Economic Literature* 56.2 (2018): 397-449.
- Lorenzoni, L., Marino, A., Morgan, D., & James, C. (2019). Health Spending Projections to 2030: New results based on a revised OECD methodology.
- Mueller, Michael T., och Alexander Frenzel. 2015. Competitive pricing within pharmaceutical classes: evidence on "follow-on" drugs in Germany 1993–2008. *The European journal of health economics* 16(1): 73-82.
- Nilsson, Jan-Eric, 2008. *Upphandling, avtalsutformning och innovationer*. Konkurrensverket
- OECD. New health technologies managing access, value and sustainability. *Managing Access, Value and Sustainability* 2017:231
- OECD, 2020, Addressing Challenges in Access to Oncology Medicines. Analytical Report
- Persson U, Svensson J, Pettersson B. A new reimbursement system for innovative pharmaceuticals combining value-based and free market pricing. *Applied Health Economics and Health Policy* 2012; 10(4):217–25.
- Stafinski T, McCabe CJ, Menon D. Funding the unfundable: Mechanisms for managing uncertainty in decisions on the introduction of new and innovative technologies into healthcare systems. *Pharmacoeconomics*, 2010;28(2):113–42.

Stokey, N. L. (1995). R&D and economic growth. *The Review of economic studies*, 62(3), 469-489

Towse, Adrian, and Elisabeth Fenwick. 2019. "Uncertainty and Cures: Discontinuation, Irreversibility, and Outcomes-Based Payments: What Is Different About a One-Off Treatment?" *Value in Health* 22 (6): 677–83. <https://doi.org/10.1016/j.jval.2019.03.013>.

Towse A och Garrison Jr LP. Can't get no satisfaction? Will pay for performance help? Toward an economic framework for understanding performance-based risk-sharing agreements for innovative medical products. *Pharmacoeconomics*, 2010;28(2):93–102.

Appendix A. ICER, Net Benefit och marginalnyttan av inkomst

När det finns många insatser som kan utföras av vården, vilka insatser ska väljas? Vi antar för enkelhetens skull att det finns olika vårdbehov i , som var och en innebär en given ökning i livskvalitet h_i och är förknippade med en konstant kostnad c_i per patient.⁴⁰ För att få så stor ökning som möjligt av livskvalitet med en given budget, ska insatser som innebär störst ökning av hälsorelaterad livskvalitet per krona väljas. Det innebär att om man rangordnar insatserna i fallande ordning från högst värde per krona till lägst, utför vi dem som har högst värde inom ramen för den budget B som är tillgänglig

$$\frac{h_1}{c_1} > \frac{h_2}{c_2} > \frac{h_3}{c_3} > \frac{h_4}{c_4}$$

Om det totala vårdbehovet för insats i är q_i har vi

$$c_1 q_1 + c_2 q_2 + c_3 \tilde{q}_3 = B$$

om de tre insatserna med högst vårdbehov ryms inom budgeten B . Vårdbehovet för insats 3 q_3 i detta exempel med lägst vårdbehov kanske inte ryms helt inom budgeten, varför $\tilde{q}_3 \leq q_3$. Insats 4 utförs inte. För enkelhetens skull antar vi här att budgeten är fix, men analysen är i princip densamma om sjukvårdsbudgeten är optimalt bestämd.⁴¹

Välfärdsökningen per krona för insats 3 har en viktig funktion. Låt R beteckna denna: $R = \frac{h_3}{c_3}$. Den motsvarar marginalnyttan av en extra krona i sjukvården. När man ska bedöma om en ny behandling n ska väljas räcker det med att jämföra med insats 3, eftersom den trängs undan. Alla insatser som har ett värde som överstiger R ska utföras av vården, andra inte. Det är också den välfärdsökning en extra krona i sjukvårdsbudget skulle ge, kostnaden i välfärd av sjukvårdsbudgetens begränsning. Behandling n är bättre än 3 om

$$\frac{h_n}{c_n} > \frac{h_3}{c_3} = R$$

Vi kan skriva om samma villkor som att insats n ska väljas om nyttan av insatsen är större än att använda resurserna c_n för insatser på marginalen.

$$h_n - R c_n > 0$$

Vi kan uttrycka samma villkor i termer av kostnader per nytta genom att bara ta inversen. Vi definierar *ICER* genom:

$$ICER_n = \frac{c_n}{h_n} < \frac{c_3}{h_3} = \frac{1}{R} = I$$

Insatsen n ska väljas framför 3 om den har ett ICER lägre än tröskelvärdet. Slutligen kan man uttrycka det i termer av *Net Benefit* (NB)

$$I h_n - c_n > 0$$

⁴⁰ Analysen kan utvidgas till att inte vara linjär.

⁴¹ För en marginell innovation följer det av enveloppeteoremet att vi kan strunta i inkomsteffekten.

Appendix B. Värdering av ATMP

Hur mycket är samhället *maximalt* villig att betala för en ny behandling? I en kostnads-effektivitetsanalys beror det på om hälso nyttan är större än att använda resurserna på andra insatser. Vi börjar med att formulera hur värdering av en ny behandling ser ut, när välfärd och kostnader är utspridda över tiden. Behandlingen kan vara kontinuerlig eller en engångsbehandling som ATMP. Värdering när behandlingsalternativ analyseras först, för att sedan modifieras till en jämförelse med en befintlig behandling.

En slutsats av analysen är att kostnader och nytta ska diskonteras olika beroende på hur tröskelvärdet utvecklas över tiden. Om vi förväntar oss att tröskelvärdet förändras över tiden ska kostnader framtida kostnader diskonteras mer än nytta.

Först ska vi härleda ett uttryck för ICER med både nytta och kostnader utspridda över tiden. I denna beräkning utgår vi från en ny behandling som inte ersätter någon befintlig. Analysen av diskontering är förenklad, för att fokusera på hur värdering av ATMP påverkas av de asymmetriska osäkerheter som diskuterats tidigare. Vi inför dessa osäkerheter sedan i följande avsnitt.

För att på enklaste sätt illustrera olika aspekter av värdering, gör vi ett antal förenklingar. Syftet är att lyfta fram principer, snarare än att vara en fullständig modell. Vi jämför en ny behandling a med en befintlig b . När det saknas alternativ behandling, antar vi att kostnad och QALY är noll. Vi utgår från att det finns två perioder, 1 och 2.

Varje behandling har kostnader och hälsoeffekt i bägge tidsperioder

- Hälsoeffekten för a är $h_a(1)$ och $h_a(2)$ i period 1 och 2.
- Kostnaden för a är $c_a(1)$ och $c_a(2)$ i period 1 och 2.

För behandling b har vi motsvarande $h_b()$ och $c_b()$. Analysen i föregående appendix kan utvidgas till att vara budgetrestriktioner vid olika tidpunkter. Motsvarande NB för respektive period kan då skrivas som

$$I(1)h_a(1) - c_a(1)$$

och

$$I(2)h_a(2) - c_a(2)$$

Tröskelvärdet för ICER kan vara annorlunda i framtiden, på grund av teknologisk innovation eller den allmänna ekonomiska utvecklingen. Eftersom vår värdering kan vara över decennier, är den också tidsberoende. Om det uppstår många nya behandlingar för olika sjukdomar med lågt ICER, kommer det framtida tröskelvärdet att sjunka. Om ett läkemedel prissätts så att ICER är nära tröskelvärdet innebär inlåsningen en framtida förlust.

Net benefit i period 1 för en insats med kostnader och QALY i bägge perioder är summan av netto nyttorna ovan, med den framtida nyttan i period 2 diskonterad med δ .

$$NB_a = (I(1)h_a(1) - c_a(1)) + \delta(I(2)h_a(2) - c_a(2)) > 0$$

För att omformulera detta till en motsvarande ICER kvott, grupperar vi kostnader och nytta per period i uttrycket för NB_a och får

$$NB_a = I(1)h_a(1) + \delta I(2)h_a(2) - (c_a(1) + \delta c_a(2)) > 0$$

Låter vi $m = I(2)/I(1)$ beteckna utvecklingen av ICER har vi

$$NB_a = I(1)(h_a(1) + m\delta h_a(2)) - (c_a(1) + \delta c_a(2)) > 0$$

Denna olikhet kan skrivas om till en ICER för behandling a jämfört med ingen behandling är givet av

$$I(1) > \frac{c_a(1) + \delta c_a(2)}{h_a(1) + m\delta h_a(2)}$$

Värdet av framtida nytta och kostnader jämförs då med tröskelvärdet i period 1. Notera att utöver den gemensamma diskonteringsfaktorn δ för kostnader, ingår också utvecklingen av tröskelvärdet m för värderingen av nytta. För vidare diskussion om olika diskonteringsfaktorer, se Claxton mfl (2011).

Fortsättningsvis kommer vi att bortse både från diskontering och förändringar i tröskelvärdet. ICER villkoret blir då

$$I > \frac{c_a(1) + c_a(2)}{h_a(1) + h_a(2)}$$

där I är tröskelvärdet för ICER i bägge perioder.

Vi har motsvarande uttryck för NB för b jämfört med ingen behandling. En fördel med NB är att uttrycket är *linjärt* snarare än en kvot som ICER. Det innebär bland annat att NB för a då b är jämförelsealternativ är skillnaden $NB_a - NB_b$. Värderingen av a jämfört med b beror på hur kostnad och hälsonytta utvecklas över tiden för respektive alternativ.

B.1. Osäkerhet

För den nya behandlingen består osäkerheten i den nya behandlingen enbart i att hälsoutfallet är osäkert. Antag att det finns en liten sannolikhet att behandlingen slutar att fungera, och det kan observeras i period 2. Om behandling och betalning kan avbrytas i period 2 blir NB bara värdet av första perioden

$$Ih_a(1) - c_a(1)$$

Om f är sannolikheten att behandlingen fortsatt verkar kommer det förväntade värdet att vara

$$E[NB_a] = (Ih_a(1) - c_a(1)) + f(Ih_a(2) - c_a(2))$$

Nyttan och kostnaden i andra perioden beror på sannolikheten f . Om nettoytan i den andra perioden är positiv är det *förväntade värdet* av behandlingen $E[NB_a]$ lägre än NB_a som den beräknas ovan. Givet att det förväntade värdet $E[NB_a]$ av behandlingen är positivt,

innebär detta att vården inte betalar mer än detta om betalning upphör om behandlingen slutar verka.⁴²

Med kontinuerlig betalning behöver inte samhället känna till sannolikheten att behandlingen slutar verka. Om däremot betalning sker i förskott måste denna sannolikhet med i beräkningarna för att inte betala för mycket. Det är en anledning till att löpande betalning för ATMP är att föredra - att betalningen kan betingas på utfall.

B.2. Det kontrafaktiska alternativet

För att beräkna värdet av den nya behandlingen a måste man fråga sig vad alternativet är. Om inte a väljs, kommer jämförelsebehandlingen b att väljas. Det innebär dock inte att det är självklart att hälsa eller pris i andra perioden kommer att vara de som gäller i första. Om inte a väljs, kommer inte all utveckling upphöra. Det finns många anledningar till att den hälsa $h_b(2)$ eller kostnad $c_b(2)$ inte motsvarar den förväntade utvecklingen. Exempelvis uppstår generisk eller terapeutisk konkurrens eller komplementära behandlingar tillkommer,

Antag att det tillkommer en ny behandling i period 2. Låt c vara behandlingen där b används i period 1 och den nya i period 2.

$$NB_c = (Ih_b(1) - c_b(1)) + (Ih_c(2) - c_c(2)) > 0$$

Eftersom behandlingen är bättre är $NB_c > NB_b$, och netto nyttan för a i förhållande till c lägre. Om vi bortser från utveckling för det kontrafaktiska alternativet överskattar vi värdet av behandling a . Alternativet till att behandla patienter med a är inte att välja behandling b för resten av patientens liv med säkerhet. Att välja b för evigt är ett värsta fall alternativ. Om bättre alternativa eller komplementära behandlingar uppstår, kommer de att väljas.

Det kontrafaktiska alternativet bör alltså vara den NB som skulle uppstå om a inte väljs, snarare än den NB som valet av b ger upphov till under resterande livstid. Den NB som kan förväntas uppstå bör tillfalla vården eller tillverkaren av ett alternativ, inte adderas till värdet av a .

Om a är en kontinuerlig behandling är detta inte ett problem. Så länge a är bättre än alternativen väljs a som behandling, om ett bättre alternativ uppstår väljs det. För ATMP behandlingar är detta ett mycket stort problem. Det är inte helt uppenbart hur en prognos för det kontrafaktiska alternativet kan utformas. Tyvärr är det en övervärdering som är svår att på ett rättvisande eller enkelt sätt korrigera för.

Problemet är som störst om läkemedelsföretaget prissätter a så att $NB_a \approx 0$, dvs ICER nära tröskelvärdet. Eftersom NB_b är en undre gräns av alternativets värde, innebär att även små osäkerheter i värderingen av alternativet att a inte bör väljas.

⁴² För behandlingar som initialt har lågt hälsovärde eller höga kostnader kan det innebära att $E[NB_a] < 0 < NB_a$. I det fallet är det optimalt att inte välja behandlingen.

B.3. Maximal och förväntad ersättning

Tröskelvärdet för ICER utgör en gräns för hur mycket samhället är villigt att betala för läkemedel. På lite sikt är det dock inte vad samhället normalt sett faktiskt betalar. Priset sjunker som sagt normalt sett över tiden med patentutgång med mera. Att samhället är villigt att betala det maximala värdet innebär inte att vi gör det i normalfallet. Särskilt om värderingen ska göras över decennier bör prisnedgångar värderas in.

Skillnaden i värdering liknar den i föregående avsnitt. Antag att vi har två olika läkemedel för livslånga sjukdomar: ett ATMP a och ett traditionellt läkemedel b . För enkelhetens skull antar vi att det saknas alternativ behandling, och vi bortser även från osäkerheter som vi diskuterat tidigare.

Oavsett om läkemedlen är alternativa behandlingar för samma sjukdom eller för olika sjukdomar, konkurrerar de om begränsade offentliga medel. Om de har samma ICER har de samma nettonyttan, $NB_a = NB_b$. Om det i period 2 kan komma en ny behandling c för patienter behandlade med b har vi

$$NB_a = NB_b < NB_c$$

Det innebär att det förväntade värdet av att behandla patienter med b i period 1 är större än att behandla (eventuellt andra) patienter med a . Avsaknaden av inlåsning gör att det förväntade värdet är högre för b - man kan byta till bättre eller billigare behandling om en sådan uppstår.

Det främsta skälet till att läkemedelspriser sjunker på lång sikt är patentutgång. Ett sätt att få behandlingsneutralitet mellan ATMP och andra läkemedel är att vid värdering av ATMP kraftigt diskontera både kostnader och QALY efter patentutgång.

Appendix C. Värdet av att vänta

Låt τ vara en förväntad prissänkning i period 2, som exempelvis uppstår genom inträde av terapeutisk konkurrens. Att vänta kan också innebära en positiv hälsoeffekt om den framtida behandlingen är bättre, alternativt negativ om att vänta på behandling innebär en degenerering. Låt $h \leq 0$ beteckna denna hälsoeffekt, och k ökningen i kvalitet av behandling i en period. Värdet av att vänta är större än att behandla i direkt om

$$\Delta(I(k+h) - (1-\tau)p) > (1+\Delta)(Ik - p)$$

där vi definierat Δ som det diskonterade nuvärdet av en enhet i framtida T perioder

$$\Delta = \sum_{t=1}^T \delta^t = \delta \frac{1 - \delta^T}{1 - \delta}$$

Det är ett värde som är växande i δ och T . När $\delta \rightarrow 1$ har vi $\Delta \rightarrow T$.

Det innebär att den framtida besparingen måste vara större än nyttoökningen i period 0.

$$\Delta(Ih + \tau p) > Ik - p = NB$$

Man kan skriva om det som hur stor den procentuella kostnadssänkningen måste vara för att motivera att vänta.

Vi kan skriva om detta i termer av net benefit och pris.

$$\tau > \frac{1}{\Delta} \frac{NB}{p} - \frac{Ih}{p}$$

Om läkemedelsföretaget föreslår en prissättning som innebär att nyttökningen är liten, kan även små framtida prissänkningar τ motivera att vänta. På motsvarande kan även en förväntan om små förbättringar i välbefinnande h i kommande behandlingar motivera att vänta med behandling. Ett liknande resonemang gäller också om det finns en liten sannolikhet att en väsentligt bättre behandling lanseras i nästa period. Även med en relativt hög diskonteringsränta på 5 procent, kan förväntade framtida prissänkningar på enstaka procent motivera att vänta, om NB är relativt litet i förhållande till behandlingens pris p .

Appendix D. Investeringsincitament

D.1. Minskad konkurrens

Antag att det finns N patienter i period 1 och $N + 1$ i period 2, där förhållandet mellan N och 1 är andelen nya patienter som tillkommer i period 2. Vi antar att patienter lever i T perioder och att sjukdomen inte ökar i befolkningen, vilket innebär att $T \leq N$. Eftersom de som behandlas med en engångsbehandling i period 1 inte behöver behandlas i period 2, är marknaden mindre i denna period med engångsbehandling än med kontinuerlig. För att jämföra vinsten av inträde tänker vi oss två företag, där företag 1 introducerar en produkt i period 1 och företag 2 introducerar en produkt i period 2. Vi jämför intäkterna för företag 2 då bägge företagen producerar botande behandlingar med intäkterna då bägge producerar kontinuerliga.

För att se att intäkterna för företag 2 med kontinuerlig behandling är större, antag att företag 2 med kontinuerlig behandling kan ta hela marknaden enbart i en period, med pris p . (Vi antar att det pris som företag 2 sätter för att få hela marknaden i båda fallen är detsamma.) Med kontinuerliga behandlingar säljer företag 2 till hela marknaden en period med en total intäkt Np . Det är större än intäkten med en engångsbehandling där företag 2 säljer till en patient över den förväntade livslängden T , eftersom $T < N$ och eftersom försäljning över tid ska diskonteras. Det förväntade värdet som en utmanare kan konkurrera om med en ATMP-tillverkare är proportionellt mot vänsterledet i följande uttryck, medan värdet som utmanaren kan konkurrera om ifall konkurrenten erbjuder kontinuerliga behandlingar är proportionellt mot högerledet:

$$\sum_{t=0}^{T-1} \delta^t p < pT \leq pN$$

Med en inträdeskostnad kommer den förväntade konkurrensen i period 2 vara lägre med engångsbehandling än med kontinuerlig behandling. Dessutom kan företaget som utvecklat en ATMP sänka priset för nya patienter utan att priset för de som redan tidigare behandlats sjunker, om ett annat företag skulle träda in.

D.2 Konkurrens och investeringsincitament

I och med att incitamenten till konkurrens minskar efter det att ATMP introduceras, visar vi här att incitamenten att vara först ökar.

Investeringen kan exempelvis modelleras som ett Tullockspel mellan två företag. Sannolikheten att företag 1 vinner tävlingen om att hinna först och därmed få patent är

$$\frac{x_1}{x_1+x_2}$$

där x_i är investeringen i FoU som företag i gör och den förväntade vinsten är

$$v \cdot \beta \cdot \frac{x_1}{x_1+x_2} + v \cdot (1 - \beta) \cdot \frac{x_2}{x_1+x_2} - x_1$$

där v är vinsten inom det nya terapeutiska området och $\beta > 1/2$ är andelen av denna som tillfaller den som hinner först med ett patent. I jämvikt gäller att

$$x_1 = x_2 = \frac{(2 \cdot \beta - 1) \cdot v}{4}$$

och vi ser att investeringen ökar i andelen som tillfaller det företag som hinner först. När $\beta = 1$ är $x_1 = v/4$ och investeringarna lämnar hälften av överskottet till företagen, en andel som givetvis sjunker när antalet företag som konkurrerar blir fler.

Tullocks modell för investeringar (eller rent-seeking behaviour) kan inte användas för att bedöma om företagen investerar för mycket eller inte. I modellen är det samhällsekonomiskt optimala att företagen investerar så litet som möjligt i FoU, eftersom sannolikheten för att investeringen ska bli framgångsrik ändå är 1.

I en mer realistisk modell kan företagen investera för mycket eller för litet, i förhållande till vad som är samhällsekonomiskt relevant, och de kan investera för tidigt eller för sent. Ibland vill vi att en större del av kakan tilldelas vinnaren; ibland vill vi dela den mer jämnt mellan ettan och de följande, beroende på hur starka incitamenten ska vara att tävla om att bedriva forskning. Litteraturen om *innovation races* fördjupar analysen. Se Stokey (1995) och efterföljande litteratur för en djupare analys.⁴³

Appendix E. Betalningar betingade på information

I en stiliserad modell kan ett ATMP antingen ha ett högt värde, H , eller lågt värde, L . Staten kan (exempelvis utifrån historisk erfarenhet) uppskatta att sannolikheten för H är S . Förväntat värde är således $SH + (1 - S)L$. Företaget har information med sannolikhet R . I så fall kan företaget avslöja information om det vill, men det kan då inte ljuga. Staten känner till R . Staten kan betala ett pris som beror på om företaget redovisar H , redovisar L eller inte redovisar någon information. Dessa priser kallas P_H , P_L och P_0 . Vi ska nu tillämpa

⁴³ Ett exempel är Denicolò, Vincenzo, and Luigi A. Franzoni. "On the winner-take-all principle in innovation races." *Journal of the European Economic Association* 8.5 (2010): 1133-1158. Lakdawalla, 2018, ger en översikt över litteraturen.

revelation principle. Företaget kan enligt våra antaganden inte felaktigt påstå H eller L , men det kan felaktigt påstå att det inte har någon information. Därför måste utfallet för företaget alltid vara minst lika bra när det väljer att avslöja H eller L , som när det inte har (eller påstår att det inte har) någon information. Vi måste sätta priserna så att det ligger i företagets intresse att tala sanning.

Antag att företaget vet H . Då kommer det att avslöja denna information om $P_H > P_0$. Antag att företaget vet L . Då kommer det att avslöja även denna information om $P_L > P_0$. Om företaget inte har någon information finns inget att avslöja och företaget får P_0 .

Notera $P_H > P_L > P_0$ strider mot värdebaserad prissättning. En sådan prissättning skulle kräva $P_H = H$ och $P_L = L$ och $P_0 = S \cdot H + (1 - S)L$. Den sista likheten strider dock mot vad som måste gälla enligt revelation principle. Företaget får vid värdebaserad prissättning mer betalt om det påstår att det inte har information, än om det påstår L . Alltså kommer det att påstå att det inte har information när det vet att värdet är L .

I vårt stiliserade exempel kan staten välja mellan olika strategier. Antingen kan det ge företaget incitament att avslöja all sin information, genom att sätta priser $P_H > P_L > P_0$, där exempelvis $P_H = H$, $P_L = L$ och $P_0 = L - \varepsilon$, vilket innebär att priset blir för lågt då företaget saknar information.

Eller så kan staten nöja sig med att ge företaget incitament att avslöja sin information då det har informationen H , exempelvis genom att sätta priset $P_H = H$ och genom att ge företaget samma priser vid L som när informationen saknas. Ett väntevärdesriktigt pris i det fallet är

$$P_0 = P_L = \frac{(1 - S)L + (1 - R)S \cdot H}{1 - R \cdot S}$$

.

Företaget får alltså ett för högt pris (i förhållande till ett värdebaserat pris) när det har informationen L och ett för lågt pris när det saknar information.⁴⁴

Poängen är att staten måste göra en avvägning mellan att ge företaget ett väntevärdesriktigt värdebaserat pris och att erbjuda priser som tvingar fram så mycket information som möjligt. Vi kan illustrera avvägningen i en något mer komplicerad modell.

Antag nu att värdet på ett ATMP är likformigt fördelat mellan 0 och 1 (exempelvis mellan 0 och 1 QALY per år under patientens återstående livstid). Liksom tidigare vet företaget med sannolikheten R precis vilket värde produkten har men med sannolikheten $1 - R$ vet det inte mer än staten. Om företaget har information om att värdet är X , där $0 < X < 1$ så kan det avslöja denna information för staten, som då kommer att vilja sätta priset $P_X = X$.

Om företaget sanningsenligt avslöjar att det saknar information är det förväntade värdet 0,5. Men antag att staten sätter $P_0 = 0,5$. Då kommer företaget att hävda att det inte har någon information, om $X < 0,5$. Det förväntade värdet, när företaget hävdar okunnighet, blir därför

$$\frac{0,5 \cdot R \cdot 0,25 + (1 - R) \cdot 0,5}{0,5 \cdot R + (1 - R)} = \frac{4 - 3R}{8 - 4R}$$

⁴⁴ Vi behöver inte anta strikta olikheter och slipper då ε , men för att förenkla framställningen använder vi strikta olikheter.

Om exempelvis $R = 2/3$ blir det förväntade värdet $3/8 = 0,375$ när företaget hävdar okunnighet. Med sannolikheten $1/3$ är företaget faktiskt okunnigt och har då en produkt vars förväntade värde är $0,5$. Med sannolikheten $1/3$ vet företaget att värdet är större än $0,5$ och kommer sanningsenligt att avslöja att $X > 0,5$. Med sannolikheten $1/3$ vet företaget att värdet är mindre än $0,5$ men det kommer att hävda okunnighet. Staten får acceptera att den betalar för mycket i en tredjedel av fallen, men korrekt värde i en tredjedel av fallen och ett värde som är väntevärdesriktigt i en tredjedel.

Om staten helt vill undvika att betala för mycket till ett företag som undanhåller information måste den erbjuda priset $P_0 = 0$ i det fall företaget hävdar okunnighet. Då kommer staten istället att betala för litet i de fall då företagen faktiskt är okunniga, det vill säga i en tredjedel av fallen.

Staten skulle också kunna välja ett sådant värde på P_0 att det förväntade värdet, i det fall företaget hävdar okunnighet, är precis lika med P_0 . Det går att visa att med de specifika parametervärdena ovan inträffar detta då $P_0 = (30.5 - 1)/2 \approx 0,36$. Företag som är genuint okunniga kommer nu att få ett pris som är lägre än det förväntade värdet, medan företag som felaktigt påstår att de är okunniga kommer att få ett för högt pris. För dessa två grupper sammantaget är dock priset lika med det förväntade värdet. För företag som hävdar ett värde $X \geq 0,36$ blir priset $P_x = X$ vilket alltså motsvarar värdet. Om sannolikheten R minskar så minskar det värde på P_0 som gör betalningen väntevärdesriktig över hela fördelningen.