

Definitief



> Geschat potentieel effectieve zorg

Schatting opbrengst focus op effectieve zorg in basispakket

Utrecht, 14 april 2021
Drian van der Woude
Jan-Peter Heida
Saskia van der Erf

SiRM. Strategies
in Regulated
Markets

Managementsamenvatting

Eén van de beleidsopties voor medisch specialistische zorg is het systematisch toetsen op effectiviteit van zorg in het basispakket. Het doel van dit rapport is om op basis van beschikbare kennis een inschatting te maken van het aandeel niet-effectieve ziekenhuiszorg in Nederland en wat er nodig is om deze niet-effectieve zorg af te bouwen.

We concluderen dat de meerkosten van huidige niet-effectieve ziekenhuiszorg naar schatting 6% van het totaal van €24,3 miljard bedragen, met een bandbreedte van 1% tot 12%. Voor deze schatting moesten veel aannames gedaan worden om negen onderzoeken naar niet-effectieve zorg te vertalen naar de Nederlandse situatie. Jaarlijks stroomt daarnaast ongeveer voor 0,3% aan niet-effectieve uitgaven in het pakket in. Geïnterviewden benadrukken het belang van effectief pakketbeheer, maar zijn terughoudend over het besparingspotentieel.

We verwachten dat circa 80% van de niet-effectieve zorg te vermijden is. Zonder wetswijzigingen gaan we ervan uit dat het met forse inspanning lukt om te stoppen met 60% van de niet-effectieve zorg. Mét wetswijzigingen denken we dat de ambitie van 80% die de HLA-partijen voor het programma Zorgevaluatie en Gepast Gebruik (ZE&GG) hebben uitgesproken, met forse inspanning haalbaar is. Voor het vermijden van 100% van de niet-effectieve zorg zouden in de hele Nederlandse zorg ideale omstandigheden gecreëerd moeten worden.

We schatten in dat met voldoende investering het merendeel van de circa 80% niet-effectieve zorg in 15 jaar afgebouwd kan worden. Het vergt circa 15 jaar om vanuit aanpassingen in wet- en regelgeving en intensivering van inspanningen niet-effectieve zorg grotendeels af te bouwen. De benodigde extra investering hiervoor loopt tijdelijk op tot circa €100 miljoen per jaar, en bedraagt structureel circa €60 miljoen per jaar.

Inhoud

Managementsamenvatting	1
1 Aanleiding en conclusie	3
2 Huidige meerkosten niet-effectieve ziekenhuiszorg zijn naar schatting 6%	4
2.1 De schatting richt zich op de meerkosten van niet-effectieve zorg	4
2.2 De meerkosten van niet-effectieve ziekenhuiszorg zijn circa 6%, met bandbreedte van 1% tot 12%	6
2.3 Niet-effectieve zorg groeit jaarlijks met ongeveer 0,3%-punt	25
2.4 Geïnterviewden benadrukken belang van effectief pakketbeheer, maar zijn terughoudend over besparingspotentieel	27
3 Circa 80% van de niet-effectieve zorg is te vermijden	28
3.1 Met forse inspanning – maar zonder wetswijzigingen – lukt het wellicht 60% van de niet-effectieve zorg te vermijden	29
3.2 Mét wetswijziging lijkt ambitie ZE&GG om 80% van niet-effectieve zorg te vermijden haalbaar	30
3.3 Voor het vermijden van 100% zijn in de hele Nederlandse zorg ideale omstandigheden nodig	31
4 Met voldoende investering kan merendeel in 15 jaar geïmplementeerd worden	32
4.1 Het vergt circa 15 jaar om vanuit wetswijziging en intensivering van inspanningen niet-effectieve zorg grotendeels af te bouwen	32
4.2 Benodigde extra uitgaven zijn tijdelijk circa €100 miljoen en structureel circa €60 miljoen per jaar	37
Bijlage 1 Bibliografie	40
Bijlage 2 Interviews en werksessies	42

I Aanleiding en conclusie

Het Centraal Plan Bureau (CPB) en de werkgroep Brede Maatschappelijke Heroverwegingen (BMH) hebben maatregelen uitgewerkt om de groei in uitgaven aan zorg te dempen. Eén daarvan betreft het toetsen van effectiviteit van zorg in het basispakket, zowel voor nieuwe behandelingen als voor behandelingen die al onder het basispakket vallen. Naast de toetsing gaat de maatregel ook uit van een strikter financieel kader en van versterking en mogelijke verbreding van toezicht en handhaving om niet-effectieve zorg zoveel mogelijk te voorkomen. Een dergelijke maatregel is ook opgenomen op de ombuigingslijst die naar verwachting een rol speelt bij de komende kabinetsformatie.

De Ministeries van VWS en Financiën hebben SiRM – Strategies in Regulated Markets - gevraagd om het besparingspotentieel van deze maatregel concreter in beeld te brengen. Ze vragen om rekening te houden met bestaande onzekerheden en mogelijke ingroeipaden aan te geven. De ministeries hebben SiRM daarnaast verzocht om te bekijken hoe de opgestelde maatregel verbeterd kan worden.

De doorlooptijd van het project was kort. We hebben in ruim één maand uitgebreid deskresearch verricht, elf interviews gevoerd en twee werksessies gehouden om de tussentijdse resultaten te toetsen en te verrijken. We hebben veel gebruikte nationale en internationale studies over niet-effectieve zorg geanalyseerd en gespiegeld aan openbare data over de Nederlandse ziekenhuiszorg. Zie Bijlage 2 voor een overzicht van geïnterviewden en deelnemers aan de werksessies.

We concluderen dat de huidige meerkosten van niet-effectieve ziekenhuiszorg naar schatting 6% zijn van de uitgaven aan medisch specialistische zorg, met een bandbreedte van 1 tot 12%. Het is belangrijk om te realiseren dat deze bevindingen sterk afhangen van vele aannames (Hoofdstuk 2). We verwachten dat met wetwijzigingen 80% hiervan daadwerkelijk te vermijden is (Hoofdstuk 3). Met aanvullende maatregelen en investeringen kan dit naar verwachting in circa 15 jaar bereikt worden (Hoofdstuk 4).

2 Huidige meerkosten niet-effectieve ziekenhuiszorg zijn naar schatting 6%

De meerkosten van huidige niet-effectieve medisch specialistische zorg (§2.1) bedragen naar schatting 6%, met bandbreedte van 1% tot 12% (§2.2). Voor deze schatting moesten veel aannames gedaan worden om negen onderzoeken naar niet-effectieve zorg te vertalen naar de Nederlandse situatie. Jaarlijks stroomt er ongeveer voor 0,3% aan niet-effectieve uitgaven in het pakket in (§2.3). Geïnterviewden benadrukken het belang van effectief pakketbeheer, maar zijn terughoudend over besparingspotentieel (§2.4).

2.1 De schatting richt zich op de meerkosten van niet-effectieve zorg

De resultaten in dit rapport vergen een zorgvuldige interpretatie. Dat gaat zowel over de afbakening van het soort zorg (§2.1.1) als de afbakening van de meerkosten (§2.1.2).

2.1.1 Dit rapport beperkt zich zo goed mogelijk tot niet-effectieve zorg

Ons onderzoek gaat in op effectief pakketbeheer. Eerste schattingen voor de budgettaire impact komen terug in de Brede Maatschappelijke Heroverwegingen (BMH), Zorgkeuzes in Kaart (ZiK) en de Ombuigingslijst van Financiën (OF).

Net als in BM, ZiK en OF, pogen we in dit rapport geen effecten mee te nemen van substitutie van zorg of andere manieren van contracteren. Enige overlap is echter onvermijdelijk, bijvoorbeeld omdat niet-effectieve zorg ook vervangen kan worden door wel effectieve zorg in de eerste lijn. In dat geval is dus sprake van zowel niet-effectieve zorg, als substitutie.

Brede Maatschappelijke Heroverwegingen scheiden niet-effectieve zorg van substitutie

In de Brede Maatschappelijke Heroverwegingen “Naar een toekomstbestendig zorgstelsel” van 2020 zijn veertig beleidsopties opgenomen om de kostengroei van de zorgsector te beteugelen. Er is gewerkt met twee varianten:

- Variant A: “Versterken gereguleerde concurrentie”
- Variant B: “Versterken gereguleerde samenwerking”

Voor beide varianten is onder het deel ‘Pakketbeheer’ met €470 tot €600 miljoen gerekend als structureel effect van beleidsoptie 6; “Versterken toetsen van zorg op (kosten)effectiviteit”. In de

uitwerking van de beleidsoptie is aangegeven dat het cumulatief effect van stringenter pakketbeheer gedurende ‘tien tot tientallen jaren’ ruim €500 miljoen is. Om dat te realiseren, gaat jaarlijks €30 miljoen aan aanvullende capaciteit naar het Zorginstituut voor bredere toetsing. Op termijn is het saldo ten minste €470 miljoen¹.

Daarnaast is onder de beleidsvariant B ‘Versterken gereguleerde samenwerking’ met de beleidsoptie 17 ‘Allocatieve normstelling’ structureel met €100 miljoen tot €200 miljoen gerekend. Het betreft hier zorg die nu in de tweede lijn geleverd wordt, maar doelmatiger in de eerste lijn geleverd kan worden. De schatting van €100 miljoen tot €200 miljoen is gebaseerd op onderzoek van het Zorginstituut op basis van 25 praktijkvoorbeelden. In ons onderzoek gaan we niet in op dergelijke substitutie om dubbeltellingen te voorkomen. Overigens wordt deze beleidsoptie niet ingeboekt in variant A.

Zorgkeuzes in Kaart noemt naast niet-effectieve zorg ook niet-kosteneffectieve zorg

Het Centraal Planbureau (CPB) bracht in 2020 ‘Zorgkeuzes in Kaart’ uit. Daarin zijn diverse voorstellen van politieke partijen doorgerekend. Voorstel 108 heet “Versterking toetsing basispakket op (kosten-) effectiviteit”. Onze opdracht richt zich op variant A1 en A2:

- A1: bewezen effectiviteit als voorwaarde voor opname in basispakket.
- A2: bewezen effectiviteit al basis voor kwaliteitsstandaarden.

Voor beide varianten zijn de opbrengsten als PM post opgenomen.

Maatregel 108 heeft mogelijk enige overlap met maatregel 109, “Gepast gebruik en zinnige zorg” waarbij wordt ingezet op betrouwbare informatie over ongewenste praktijkvariatie. Deze maatregel, waaronder een openbaar register waarin wordt opgenomen wat redelijk te verwachten volumes voor een ingreep zijn, kan bijdragen aan het realiseren van effectieve zorg.

Ombuigings- en Intensiveringslijst Financiën verwijst naar Zorgkeuzes in Kaart

Op de ombuigings- en intensiveringslijst van Financiën is de impact van honderden maatregelen doorgerekend. Voor VWS luidt maatregel 18 “Bewezen effectiviteit van zorg als voorwaarde voor opname in het basispakket”. Deze maatregel kost jaarlijks €120 miljoen. Voor de onderbouwing wordt verwezen naar Zorgkeuzes in Kaart. In de lijst zijn de opbrengsten nog niet opgenomen.

2.1.2 Meerkosten leiden niet altijd direct tot besparingspotentieel

In dit rapport spreken we over meerkosten van niet-effectieve zorg. Dit is het saldo van te vermijden huidige uitgaven aan niet-effectieve zorg en uitgaven aan het alternatief.

In sommige gevallen is het alternatief géén zorg, zodat de besparing gelijk is aan de huidige uitgaven aan de niet-effectieve zorg. In andere gevallen wordt de niet-effectieve zorg vervangen

¹ Het fiche waar de maatregel 6 verder in is uitgewerkt, geeft ook nog aan dat €128 miljoen per jaar bespaard zou kunnen worden door de grens voor een QALY te verlagen van €80.000 naar €70.000. Deze verlaging nemen wij niet mee in ons onderzoek.

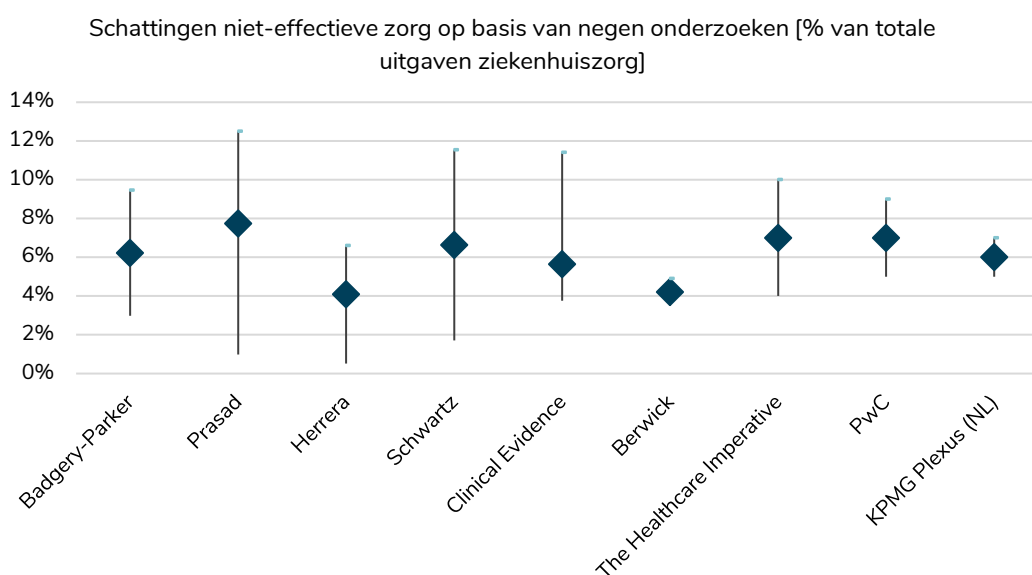
door een goedkoper alternatief, zodat slechts een deel van de uitgaven aan niet-effectieve zorg wegvallen door effectief pakketbeheer. Hiervoor corrigeren we in onze schattingen voor het aandeel van niet-effectieve zorg.

We benadrukken dat meerkosten zich niet altijd direct vertalen in besparingspotentieel. Door minder zorg te leveren, valt slechts een deel van de kosten direct weg. Vaste kosten, zoals een diagnostisch laboratorium of personeelskosten, kunnen alleen met organisatorische veranderingen verlaagd worden. Het kan bijvoorbeeld betekenen dat laboratoria moeten fuseren als minder diagnostiek plaatsvindt. Dergelijke (transmurale) organisatorische veranderingen vergen doorgaans langdurige trajecten. Maar op middellange termijn zijn deze kosten wel aan te passen.

Minder uitvoeren van niet-effectieve zorg binnen plafondafspraken creëert ruimte voor andere (effectieve) zorg. Het besparingspotentieel kan in onze ogen het best in die context gezien worden: het draagt bij aan de ambitie van HLA-partijen om de groei aan zorguitgaven te dempen.

2.2 De meerkosten van niet-effectieve ziekenhuiszorg zijn circa 6%, met bandbreedte van 1% tot 12%

We concluderen dat de meerkosten van niet-effectieve medisch specialistische zorg (msz) in Nederland circa 6% bedragen van de totale jaarlijkse uitgaven van €24,3 miljard (Figuur 1), met een bandbreedte tussen 1% en 12%. Deze schatting is tot stand gekomen door de bevindingen uit onderzoeken naar niet-effectieve zorg te vertalen naar de Nederlandse medisch specialistische zorg. We gebruiken daarvoor een rekenmodel dat expliciet rekening houdt met de grote hoeveelheid onzekere factoren die bij deze berekening onvermijdelijk een rol spelen. Hieronder lichten we de methode en gebruikte aannames steeds zo transparant mogelijk toe.



Figuur 1 Schatting huidige meerkosten niet-effectieve msz op basis van vertaling negen onderzoeken naar Nederlandse praktijk.

2.2.1 Schattingen kennen meerdere onzekere factoren waarvoor scenario's zijn opgenomen

Om de resultaten van onderzoeken te vertalen naar het aandeel niet-effectieve msz in Nederland zijn steeds meerdere aannames nodig. Uitzondering hierop in Figuur 1 is de studie van KPMG Plexus, waarin wel een schatting voor Nederland op totaalniveau wordt gepresenteerd, maar ook in die publicatie zelf worden meerdere aannames gedaan om resultaten uit deelgebieden op te schalen naar de totale msz.

Deze vertaalslagen hebben we samengevat in een rekenmodel dat we op elk onderzoek toepassen (Figuur 2). Het startpunt van het rekenmodel is steeds een getal uit het specifieke onderzoek waarna we systematisch aannames doen. De stappen zijn achtereenvolgens:

- 1 Factor Nederland. Gevonden resultaten in het buitenland zijn niet direct te vertalen naar Nederland. In Nederland wordt minder zorg per hoofd van de bevolking geleverd.
- 2 Factor opschaling vanuit deelgebieden. Vaak geeft een onderzoek een indicatie van het aandeel niet-effectieve zorg in een deelgebied van de zorg, bijvoorbeeld voor diagnostiek. Dit is niet direct op te schalen naar de totale msz.
- 3 Factor aandeel zorg met onbekende effectiviteit. Tenzij hier een expliciete uitspraak over gedaan wordt, kan onderzoek alleen toegepast worden op het deel van de zorg met nog onbekende effectiviteit.
- 4 Factor aandeel niet-effectieve zorg binnen het deel met onbekende effectiviteit. Tenzij hier een expliciete uitspraak over gedaan wordt, moet een aanname gedaan worden over het aandeel niet-effectieve zorg binnen de zorg met onbekende effectiviteit.
- 5 Factor meerkosten. Als het onderzoek gaat over niet-effectieve zorg die vervangen wordt door goedkopere, wel-effectieve zorg, kunnen niet alle kosten van de niet-effectieve zorg in mindering gebracht worden.
- 6 Factor differentiatie. Niet-effectieve zorg is vaak gedifferentieerd: voor een deel van de patiënten is het wel effectief en voor een ander deel niet. Als dit onderscheid door behandelaren voorafgaand aan een behandeling niet eenduidig te maken is, kan redelijkerwijs niet alle (retrospectief) niet-effectieve zorg (prospectief) voorkomen worden.



Figuur 2 In het rekenmodel wordt de input van een specifiek onderzoek in zes stappen vertaald naar de meerkosten van niet-effectieve Nederlandse msz.

Als uit het specifieke onderzoek niet duidelijk is welke waarde voor bovenstaande factoren gekozen kan worden, maken we steeds gebruik van een laag, midden en hoog scenario. Hieronder lichten we toe welke waardes we voor die scenario's hanteren en wat de argumentatie daarvoor is.

Scenario's factor Nederland

Voor de vertaling van internationaal onderzoek naar de Nederlandse praktijk maken we gebruik van (Tabel 1):

- Een vermenigvuldigingsfactor 50% voor onderzoek naar het voorkomen van niet-effectieve zorg in de VS (scenario laag).
- Een factor 66% voor onderzoek over overige Westerse landen (scenario midden).
- Een factor 100% voor onderzoek over Nederland (scenario hoog).

Scenario	Waarde	Toelichting
Laag	50%	Het lage scenario wordt toegepast op onderzoek uit over VS
Midden	66%	Het midden scenario wordt toegepast op onderzoek over overige Westerse landen
Hoog	100%	Het hoge scenario wordt toegepast op onderzoek over Nederland of het VK

Tabel 1 Voor de vertaling van onderzoek naar niet-effectieve zorg naar Nederland houden we rekening met een laag (50%) en midden (66%) scenario.

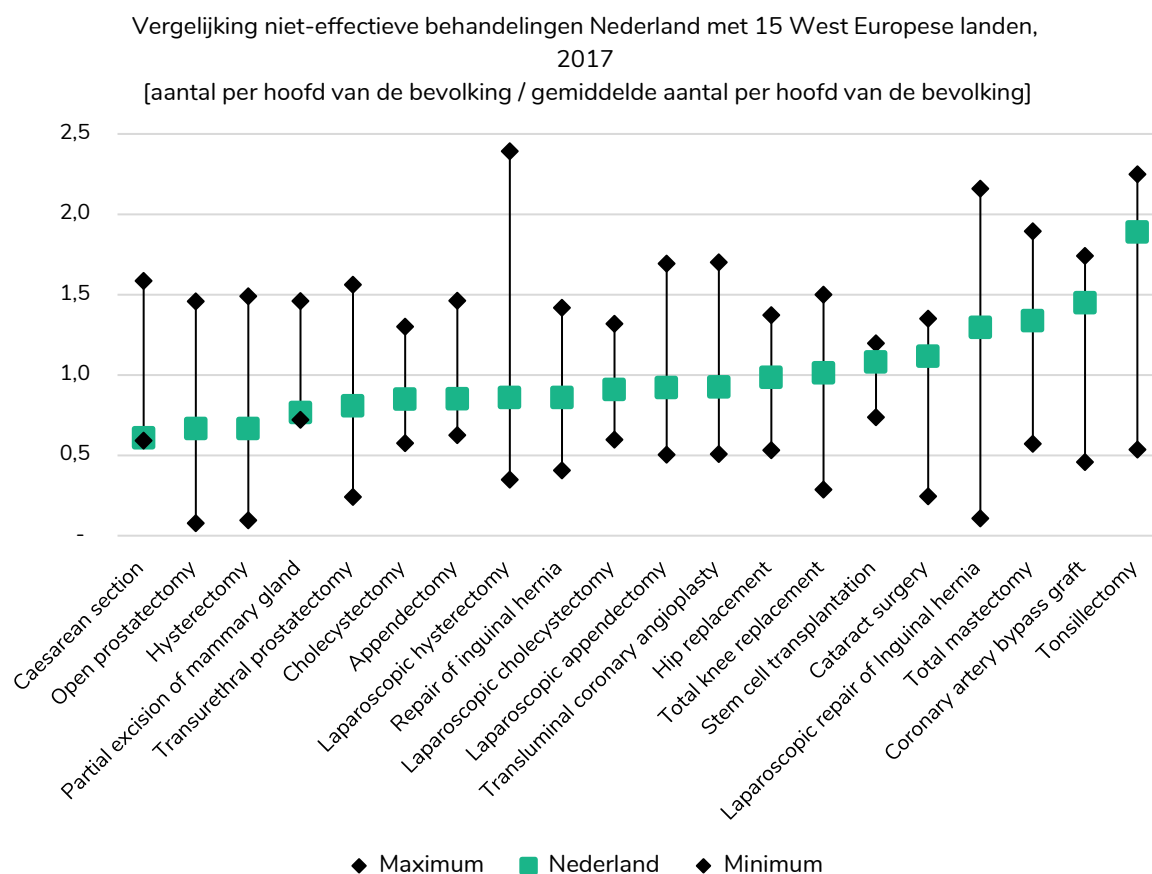
De VS staan er om bekend dat er door een andere gezondheidscultuur, met veel 'defensive medicine', meer niet-effectieve zorg wordt geleverd. We doen in dit rapport de grove aanname dat niet-effectieve zorg in Nederland half zo vaak voorkomt als in de VS. Volgens cijfers van de OESO worden in de VS bijvoorbeeld 255 CT-scans per hoofd van de bevolking gedaan, tegen 96 in Nederland. Voor MRI-scans is dat 110 vs. 51 (Figuur 4). Onderzoek van PwC schat dat defensive medicine ongeveer 10% van de uitgaven aan msz (PwC Health Research Institute, 2010). Een deel daarvan zal niet-effectieve zorg zijn.

Ook ten opzichte van vijftien vergelijkbare West-Europese landen wordt in Nederland relatief weinig zorg per inwoner geleverd. Dat geldt zowel voor een selectie van veelvoorkomende behandelingen (Figuur 3), diagnostiek (Figuur 4) als geneesmiddelen (Figuur 5, Figuur 6).

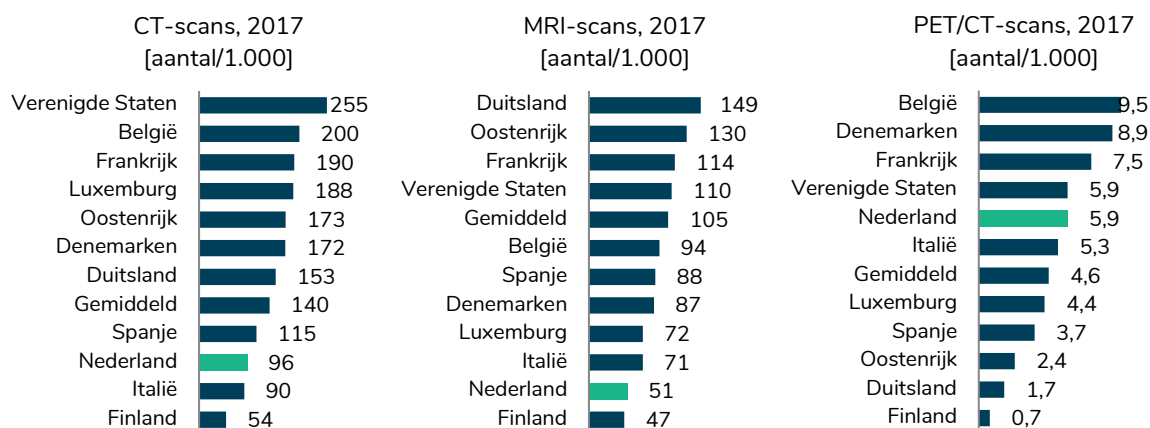
Voor vergelijkingen met het Verenigd Koninkrijk gebruiken we het scenario hoog (100%). Daar wordt door het National Institute of Clinical Excellence (NICE) veel onderzoek gedaan voor zorg die door de National Health Service (NHS) wordt geleverd. Daarbij worden behandelingen eerder als niet (kosten)effectief aangemerkt als in Nederland². Voor 17 van de 20 ingrepen waarvoor de cijfers presenteert (Figuur 3), worden er in Engeland gemiddeld circa een derde minder gedaan dan in Nederland³.

² In Nederland is de grens circa €80.000 / Qaly. NICE gebruikt >£30.000 ofwel €35.000 / Qaly waarbij overigens wel aparte fondsen zijn voor oncologische behandelingen zodat de de-facto grens hoger is.

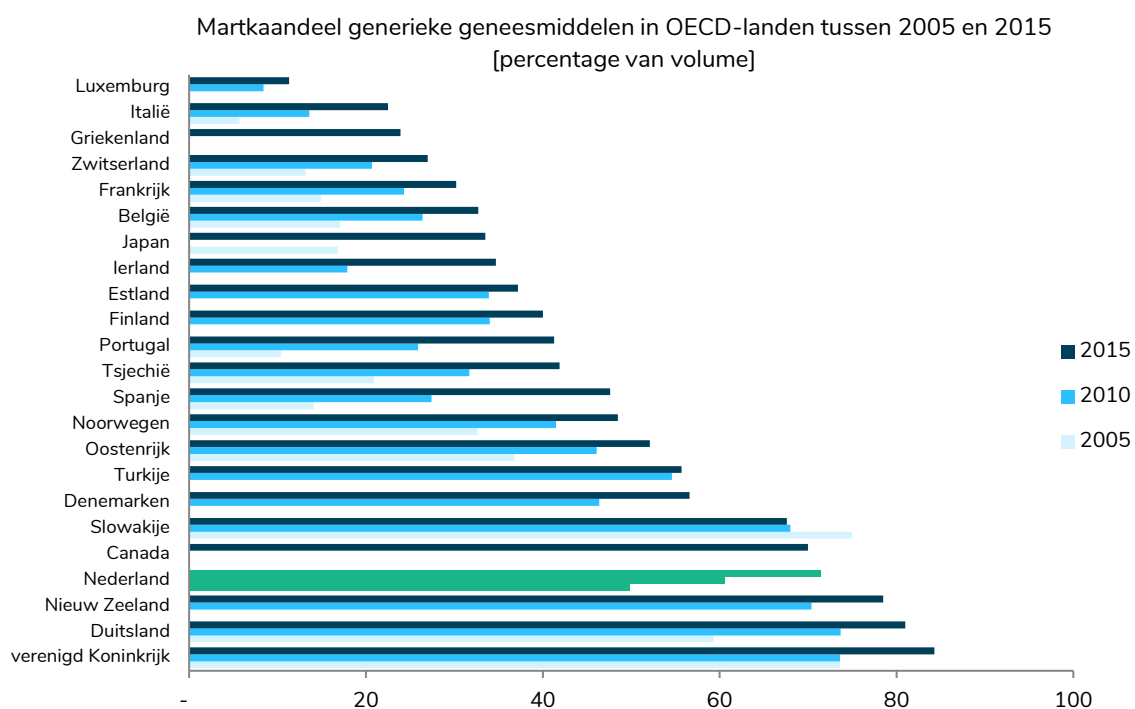
³ De uitzonderingen zijn keizersnede, hysterectomie, en gedeeltelijke borstverwijdering.



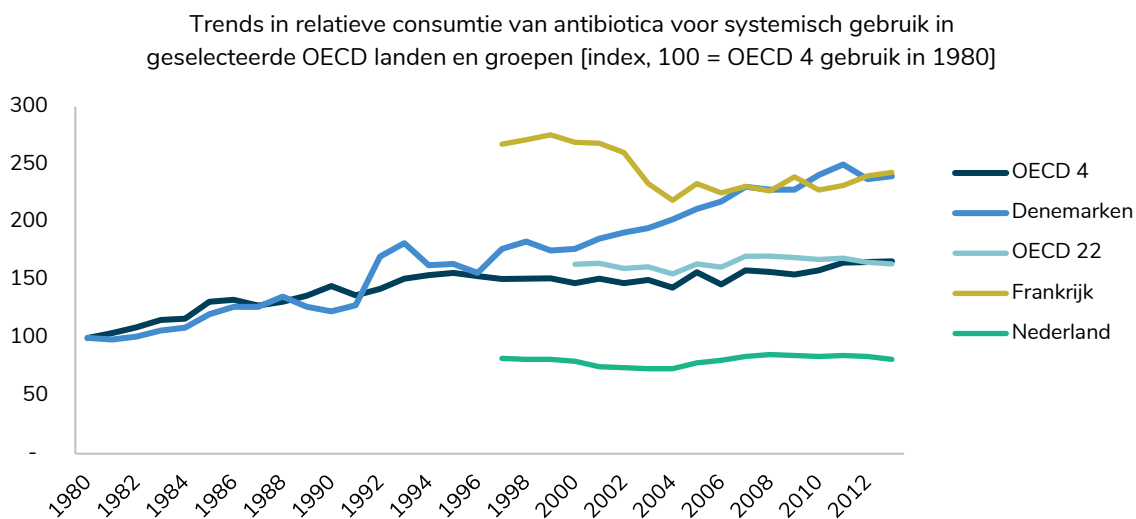
Figuur 3 Nederland scoort relatief weinig niet-effectieve behandelingen vergeleken met andere Europese landen. Bron: OECD.



Figuur 4 Nederland voert relatief weinig diagnostiek uit. Bron: OECD.



Figuur 5 Nederland maakt relatief veel gebruik van generieke geneesmiddelen. Bron: OECD Health statistics (2016).



Figuur 6 Nederland maakt relatief weinig gebruik van antibiotica. Bron: OECD Health statistics (2016).

Scenario's factor opschaling

Voor de vertaling van onderzoek in deelgebieden naar de totale msz maken we gebruik van opschalingsfactoren. De meest relevante factoren zijn:

- Een factor 0,7 voor onderzoek naar het aandeel niet-effectieve zorg die te classificeren is als activiteit diagnostiek.
- Een factor 0,7 voor onderzoek naar ingrepen die te classificeren zijn als soort zorg niet-complex veelvoorkomend.

- Een factor 1,0 voor onderzoek dat over een deelgebied dat te classificeren is als activiteit operatieve verrichting.

Deze factoren zijn tot stand gekomen door steeds het potentieel en het kostenaandeel van een deelgebied te wegen (Tabel 2). Het potentieel van een deelgebied geven we steeds een factor 1,0 (hoog), 0,75 (midden), of 0,5 (laag). De inschatting is daarbij dat in een 'hoog' deelgebied twee keer zoveel niet-effectieve zorg voorkomt als in een 'laag' deelgebied. Dit wordt vervolgens gewogen met het aandeel in de totale ziekenhuiskosten van de verschillende deelgebieden om tot de uiteindelijke opschalingsfactor te komen.

Doorsnede	Categorie	Aandeel kosten	Relatief gewicht	Factor opschaling
Activiteit	Beeldvorming en diagnostiek	15%	1,00	0,7
	Geneesmiddelen	10%	0,75	1,0
	Kliniek	30%	0,50	1,3
	Operatieve verrichtingen	15%	0,75	1,0
	Polikliniek en SEH-bezoek	30%	0,75	1,0
Type zorgproduct	Chronisch	39%	1,00	0,9
	Electief	33%	1,00	0,9
	Acuut	16%	0,75	1,2
	Oncologie	9%	0,50	1,8
	Overig	3%	0,50	1,8
Soort zorg	Complex	20%	0,50	1,4
	Veelvoorkomend niet-complex	35%	1,00	0,7
	Minder voorkomend niet-complex	35%	0,75	1,0
	Met geneesmiddelen	10%	0,75	1,0

Tabel 2 Voor de opschaling van onderzoek over een deelgebied van de msz, maken we steeds gebruik van één van een indeling naar activiteit, type zorgproduct of soort zorg en wegen we met het kostenaandeel en geschat potentieel om tot de opschalingsfactor te komen.

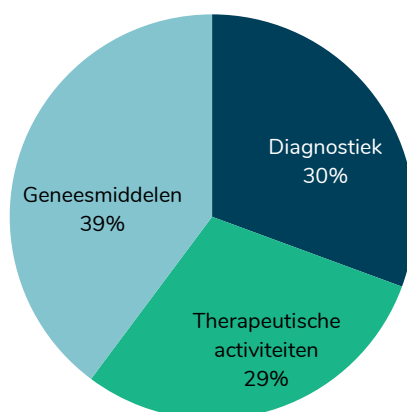
Uit een analyse van op dit moment bekende voorbeelden van niet-effectieve zorg en de gesprekken die we hierover hebben gevoerd, concluderen we dat niet-effectieve zorg vaker voorkomt binnen beeldvorming en diagnostiek en veelvoorkomende niet-complexe zorg dan in andere deelgebieden.

Bekende voorbeelden van niet-effectieve zorg vinden we vooral voor het gebruik van geneesmiddelen, diagnostiek en laag-complexe veelvoorkomende therapeutische activiteiten. De oververtegenwoordiging van diagnostiek en geneesmiddelen volgt bijvoorbeeld uit analyse van 1366 'beter niet doen' activiteiten in Nederland (Figuur 7), (Wammes & al., 2016). Een verdere analyse van de Verstandige Keuzes lijst van de FMS die mede hieruit is voortgekomen, leert dat het bij therapeutische activiteiten veelal om laag-complexe, veel voorkomende activiteiten gaat. Ook de 'Choosing Wisely' lijsten uit de VS, Canada, Zwitserland en Frankrijk, kennen verhoudingsgewijs veel van dergelijke behandelingen.

Uit gesprekken concluderen we dat bovenstaand beeld hoogstwaarschijnlijk niet alleen voor het 'laaghangend fruit' geldt. Voor complexe zorg of aandoeningen met hoge ziektelast zoals oncologie, is de effectiviteit van de geleverde zorg over het algemeen meer aan onderzoek onderhevig. Uiteraard is er wel discussie over (kosten)effectiviteit van complexe, dure behandelingen. Bijvoorbeeld bij grote investeringen in protontherapie of inzet van nieuwe robottechnologie (CPB, 2017). Zorgverzekeraars proberen daar ook de kosten te beheersen, door maar bij enkele partijen protontherapie in te kopen, of door dezelfde prijs te betalen voor behandelingen met en zonder robot.

Ook geldt voor bepaalde vormen van zorg, zoals acute zorg, dat RCT's minder geschikt zijn. Als we kijken naar uitgaven aan de (poli)kliniek, is de inschatting dat hier wel doelmatigheidswinst mogelijk is, maar dat het hierbij minder om niet-effectieve zorg gaat. Alleen als een opname of polikliniekbezoek een direct gevolg is van een niet-effectieve behandeling heeft het invloed op die kosten. Tot slot stellen we het aandeel niet-effectieve zorg binnen de geneesmiddelen qua kosten aandeel naar beneden bij, omdat voor dure geneesmiddelen al een sluis-mechanisme in Nederland bestaat.

Analyse 1366 'beter niet doen' activiteiten in NL naar soort verrichting
[ongewogen telling]



Figuur 7 Beter niet doen activiteiten betreffen relatief veel diagnostiek en geneesmiddelen. Bron: Wammes et al. (2016).

Het gehanteerde kostenaandeel is een schatting op basis van projectervaring van SiRM. In werkelijkheid zullen de percentages hier licht van afwijken, maar deze onzekerheid valt binnen de gevoeligheid van de analyse in dit rapport.

Scenario's factor niet-bewezen

Als in een onderzoek geen uitspraak wordt gedaan over het aandeel niet-bewezen zorg, hanteren we een laag (25%), midden (33%) en hoog (40%) scenario (Tabel 3).

Scenario	Waarde	Toelichting
Laag	25%	Op basis van enquête onder FMS leden en onderzoek Imrie & Ramey
Midden	33%	Midden van hoog en laag
Hoog	40%	Op basis van analyse richtlijnen gynaecologie, neurologie en orthopedie.

Tabel 3 Voor de factor niet bewezen houden we, tenzij expliciet anders aangegeven, rekening met scenario's tussen 25% en 40%.

Het lage scenario is gebaseerd op een enquête onder medisch specialisten (Eén vandaag opiniepanel, 2015) en onderzoek van (Imrie & Ramey, 2000) naar niet-effectieve zorg. Uit de enquête is af te leiden dat medisch specialisten in grofweg 25% van de gevallen aangeven niet te weten welke behandeling, wetenschappelijk gezien, de beste keuze voor de patiënt is. Hiervoor is een gewogen gemiddelde genomen van de resultaten uit Figuur 8, waarbij de categorie geen mening / weet niet buiten beschouwing is gelaten. Imrie & Ramey schatten op basis van een inventarisatie van 2000 bekende onderzoeken naar de mate van bewijslast voor medische behandelingen dat 76% van de zorg een bepaalde vorm van overtuigend bewijs kent, waarvan de helft in de vorm van Randomized Control Trials (RCT's). Ze geven daarbij aan dat RCT's niet altijd geschikt zijn en het dus redelijk kan zijn naar andere vormen van bewijs te kijken.



Figuur 8 Medisch specialisten geven aan dat zij gemiddeld in 25% van de gevallen niet weten wat wetenschappelijk gezien de beste behandeling is voor de patiënt. Bron: Eén vandaag opiniepanel (2015).

Het hoge scenario is gebaseerd op een interpretatie van onderzoek naar het bewijsniveau in Nederlandse richtlijnen voor gynaecologie, neurologie en orthopedie (FMS, 2016). Daaruit blijkt dat ongeveer de helft van de medische handelingen bewijsniveau 3 of 4 op een schaal van 4 heeft, wat over het algemeen als onvoldoende wordt beschouwd. Richtlijnen gaan vooral in op hoe een behandeling moet worden uitgevoerd en niet zozeer of de behandeling moet worden uitgevoerd. Verder is ook dit aandeel tot stand gekomen op basis van een kale telling en niet gewogen naar patiënten of kosten. We stellen de factor een half daarom iets naar beneden bij en komen uit op een hoog scenario van 40% onbewezen zorg. Dit sluit ook aan bij het inzicht dat de resultaten van de Clinical Evidence studie uit het VK naar beneden bijgesteld moeten worden (zie paragraaf §2.2.2).

Scenario's factor niet-effectief binnen niet bewezen

Als in een geselecteerd onderzoek geen uitspraak wordt gedaan over het aandeel niet-effectieve zorg binnen niet-bewezen zorg, hanteren we een laag (20%), midden (30%) en hoog (40%) scenario (Tabel 4).

Scenario	Waarde	Toelichting
Laag	20%	Op basis van onderzoek Herrera-Perez et al. (2018)
Midden	30%	Midden van hoog en laag Vertaling analyse van 200 onderzoeken uit ZonMw Doelmatigheidsonderzoek programma
Hoog	40%	Op basis van onderzoek Prasad et al. (2013)

Tabel 4 Voor de factor niet-effectief binnen niet bewezen houden we, tenzij expliciet anders aangegeven, rekening met scenario's tussen 20% en 40%.

De onderzoeken van Herrera-Perez en Prasad worden hier niet opgevoerd voor een opschaling van deze onderzoeken zelf (zie daarvoor §2.2.2) We gebruiken deze onderzoeken hier voor andere onderzoeken waarin geen uitspraak over het aandeel niet-effectief binnen niet-bewezen wordt gedaan.

We hebben de waardes in Tabel 4 niet direct uit de publicaties overgenomen. De resultaten uit Prasad suggereren op het eerste gezicht dat circa 50% van de onderzoeken naar bestaande zorg aantonen dat er geen meerwaarde is (zie §2.2.2). Het betreft echter onderzoek in een vakblad uit de VS, wat voor ons reden is dit percentage naar beneden bij te stellen. Tegelijkertijd betreft het een review van historische publicaties, die niet direct op zoek zijn naar niet-effectieve zorg. Bij de publicaties over bestaande verrichtingen gaat het dan ook vrijwel altijd om 'medical reversals', waarbij een andere behandeling beter is. Deze effecten compenseren we grofweg door het percentage bij te stellen naar 40% (volgens midden scenario in §2.2.2). Bij Herrera-Perez geldt dezelfde overweging, met uitzondering van het feit dat in dat onderzoek een mix van publicaties uit de VS en Europa is bekeken. Vandaar dat we de conclusie uit die studie van 20% hier wel overnemen (zie midden scenario Herrera-Perez in §2.2.2).

Een analyse van 200 onderzoeken van het ZonMw Doelmatigheidsprogramma laat een vergelijkbaar beeld zien. Ongeveer de helft van de 200 onderzochte zorgactiviteiten toonde zorginhoudelijk noch financieel meerwaarde aan ten opzichte van andere behandelingen. In een reactie per mail geeft ZonMw echter aan dat deze onderzoeken niet representatief zijn voor de totale zorg met onvoldoende wetenschappelijk bewijs. Ook is het onderzoek gericht op doelmatigheid en betreft het dus meer dan niet-effectieve zorg volgens onze definitie. Het lijkt daarom redelijk dit bij te stellen tot de scenario's 20 tot 40%.

Scenario's factor meerkosten

Het wegvallen van niet-effectieve zorg leidt niet altijd tot het volledig wegvallen van de zorgkosten voor de patiënt. Dit geldt in het bijzonder voor studies waarbij sprake is van 'medical reversals', waarbij expliciet sprake is van een vervangende behandeling. In dat geval nemen we aan dat maar een kwart van de kosten wegvalt door de niet-effectieve zorg te deïmplementeren.

Als expliciet wel sprake is van zorg die in het geheel overbodig is en ook niet vervangen wordt door andere zorg, vallen wel alle kosten weg. Dat leidt tot de scenario's uit Tabel 5.

Scenario	Waarde	Toelichting
Laag	25%	Bij medical reversals is sprake van een vervangende behandeling en valt dus maar een deel van de kosten weg
Midden	66%	Midden van hoog en laag
Hoog	100%	Als geen sprake is van een vervangende behandeling vallen alle kosten weg.

Tabel 5 Voor de factor meerkosten houden we rekening met scenario's tussen 25% en 100%.

Scenario's factor differentiatie

Of iets niet-effectieve zorg betreft, is vaak van tevoren maar gedeeltelijk vast te stellen. Het gaat vrijwel nooit om behandelingen die in het geheel niet effectief zijn – het gaat bijna altijd om niet-effectiviteit voor een specifieke subgroep. In sommige gevallen is desalniettemin goed van tevoren te bepalen of het om effectieve zorg gaat, bijvoorbeeld bij een duidelijke leeftijdsgrens. In andere gevallen is het minder eenduidig, als bijvoorbeeld de mate van pijn de indicatiestelling mede bepaalt. In het rapport werken we hiervoor met drie scenario's voor een correctiefactor, uiteenlopend van 25% (laag) tot 100% (hoog), zie Tabel 6.

Scenario	Waarde	Toelichting
Laag	25%	Indicatiestelling is enigszins mogelijk: het is lastig vooraf te bepalen hoe de patiënt zal reageren op de medische verrichting. Niet-effectiviteit is alleen ex-post goed aan te tonen.
Midden	66%	Midden van hoog en laag
Hoog	100%	Perfekte indicatiestelling mogelijk: voor alle patiënten is vóóraf te bepalen of de medische verrichting inderdaad effectief is of niet.

Tabel 6 Voor de factor differentiatie houden we rekening met scenario's tussen 25% en 100%.

2.2.2 Opschaling diverse studies leidt met veel aannames tot circa 6% niet-effectieve zorg

De negen in Figuur 1 genoemde studies zijn allemaal aan de hand van het hierboven toegelichte rekenmodel vertaald naar een schatting voor de meerkosten van niet-effectieve msz in Nederland. Hieronder lichten we per onderzoek toe hoe we hiertoe gekomen zijn.

Badgery-Parker: 3 tot 10%

De benadering van Badgery-Parker et al. is anders dan veel andere studies (Badgery-Parker & al., 2018). In deze benadering kijken ze namelijk naar een subgroep van 27 verrichtingen in Australië, waarvan bekend is dat ze bij een deel van de populatie niet effectief zijn, maar bij een ander deel wel. Ze publiceren vervolgens het aandeel van de totale uitgaven aan deze 27 verrichtingen dat niet-effectief is. Ze vonden dat 7,4%-14,7% van de kosten voor deze ingrepen retrospectief te

beoordelen zijn als van lage waarde, afhankelijk van hoe ruim de populatie waarvoor deze zorg van lage waarde is wordt gedefinieerd. Dit kent onzekerheid, omdat het om zorg met gedifferentieerde effectiviteit gaat, waardoor het voor de onderzoekers niet altijd volledig duidelijk of een specifieke patiënt wel of niet in aanmerking zou moeten komen voor deze zorg. Verder betreft het voornamelijk chirurgische ingrepen uit Choosing Wisely lijsten.

De inputwaarde van 7,4% tot 14,7% wordt naar beneden bijgesteld tot 3% tot 9,5% na toepassing van het rekenmodel (Tabel 7). De belangrijkste reden daarvoor is dat het onderzoek uit het buitenland betreft en dat het gaat om 'low-value care' dat ook zorg omvat waarvoor een vervangende behandeling de voorkeur heeft, waardoor waarschijnlijk niet alle gerapporteerde kosten wegvallen. We houden daarom alleen in het scenario hoog rekening met het volledig wegvallen van de kosten. We passen op dit onderzoek geen additionele factoren niet-bewezen of niet-effectief toe. De reden daarvoor is dat we als uitgangspunt aannemen dat het gevonden percentage direct iets zegt over het aandeel niet-effectieve zorg in de totale zorg. Anders gezegd nemen we aan dat het aandeel niet-effectieve zorg binnen deze 27 verrichtingen representatief is voor het aandeel niet-effectieve zorg binnen alle operatieve verrichtingen (wat op zijn beurt weer representatief is voor de totale zorg, zie de factor 1,0). Tot slot merken we op dat de factor differentiatie door de auteurs zelf al is meegewogen door een laag en een hoog scenario te nemen.

Scenario	Input	Factor Nederland	Factor opschaaling	Factor niet-bewezen	Factor niet-effectief	Factor kosten	Factor differentiatie	Uitkomst
Laag	7,4%	66%	1,0	n.v.t.	n.v.t.	63%	n.v.t.	3,0%
Midden	11,1%	66%	1,0	n.v.t.	n.v.t.	63%	n.v.t.	6,2%
Hoog	14,7%	66%	1,0	n.v.t.	n.v.t.	100%	n.v.t.	9,5%
Toelichting		Midden	Operatieve verrichtingen			Midden tot hoog		

Tabel 7 De input uit Badgery-Parker verlagen we met een factor 0,4 (=66%*63%).

Prasad: 1 tot 13%

In 2013 inventariseerden Prasad et al. 363 artikelen uit de New England Journal of Medicine tussen 2001-2010 met onderzoek naar bestaande medische verrichtingen (Prasad, et al., 2013):

- Voor 40% van de artikelen (146) was de conclusie dat de onderzochte verrichting geen meerwaarde heeft ten opzichte van andere verrichtingen.
- Nog eens 40% vond dat de onderzochte verrichting meerwaarde had.
- Voor 20% kon geen conclusie getrokken worden.

Grofweg schatten we op basis hiervan in dat 50% van de onderzochte bestaande verrichtingen niet effectief bleken, en bestaande richtlijnen tegenspraken. Overigens onderzochten meerdere artikelen dezelfde verrichting. De auteurs geven aan dat er voorbeelden van niet effectieve zorg waren voor alle medisch specialismen.

In het model vertalen we dit naar een schatting van tussen de 1,0% en 12,5%. De belangrijkste aanpassingen zijn de vertaling van de VS naar Nederland (50%) en een opschaling voor het feit dat de onderzochte publicaties niet per se op zoek waren naar niet-effectieve zorg (Tabel 8). We nemen aan dat dit mogelijk leidt tot een onderschatting van maximaal een factor 2 (scenario hoog). Voor de factor kosten nemen we aan dat in ieder geval niet de volledige kosten wegvallen, omdat het om medical reversals gaat en op de factor differentiatie gaan we ervan uit dat de behandelingen op basis van het onderzoek grotendeels een duidelijke indicatiestelling hebben.

Een kanttekening bij deze resultaten is dat de input een telling van publicaties betreft en niet gewogen is naar patiëntenaantallen of kosten.

Scenario	Input	Factor Nederland	Factor opschaling	Factor niet-bewezen	Factor niet-effectief	Factor kosten	Factor differentiatie	Uitkomst
Laag	50%	50%	1,0	25%	n.v.t.	25%	63%	1,0%
Midden	50%	50%	1,5	33%	n.v.t.	63%	100%	7,7%
Hoog	50%	50%	2,0	40%	n.v.t.	63%	100%	12,5%
Toelichting		Laag	Alleen medical reversals bekeken	Laag tot hoog		Laag tot midden	Midden tot hoog	

Tabel 8 De input uit Prasad is een waarde voor niet-effectief binnen niet bewezen en kent vertaalslagen voor de overige factoren.

Herrera-Perez: 0,5 tot 6,5%

Herrera-Perez et al. onderzochten meer dan 3.000 randomized control trials (RCTs) die zijn gepubliceerd tussen 2003 en 2017 in drie belangrijke medisch wetenschappelijke tijdschriften (Herrera-Perez & al., 2019). Dat zijn the Journal of the American Medical Association, the Lancet, en the New England Journal of Medicine .

We concluderen dat 20% van de 1650 RCT's naar bestaande verrichting een medical reversal betreft. Ze vonden bijna 400 'medical reversals', waarvan het merendeel ook besproken in een systematische review. Op basis daarvan schatten we dat ongeveer drie kwart inderdaad een medical reversal betreft. Uit de systematische reviews bleek:

- Ruim de helft (53%) betreft inderdaad een medical reversal.
- Voor bijna een derde (28%) is het niet duidelijk of het inderdaad een medical reversal is.
- Een vijfde niet is opgenomen in een systematische review.

Van de laatste twee categorieën gaan we ervan uit dat de helft daarvan ook een medical reversal betreft. Dat brengt ons op ongeveer 300 'medical reversals'. Dit is iets minder dan 20% van de bijna 1.650 onderzochte RCTs van bestaande verrichtingen die in dezelfde periode zijn gepubliceerd in deze drie medisch wetenschappelijke bladen. Nieuwe verrichtingen (of geneesmiddelen) die enkel nog maar in een research setting worden toegepast, zijn dus niet meegenomen.

In het model vertalen we dit naar een schatting van tussen de 0,5% en 6,6% (Tabel 9). Hiervoor gebruiken we grotendeels dezelfde stappen als bij de analyse van Prasad, met de uitzondering dat het in deze studie om een mix van onderzoek uit de VS en Europa gaat, waardoor we met het midden-scenario werken voor de vertaling naar Nederland.

Ook bij dit onderzoek is een kanttekening dat de input een telling van publicaties betreft en niet gewogen is naar patiëntenaantallen of kosten.

Scenario	Input	Factor Nederland	Factor opscha-ling	Factor niet-bewezen	Factor niet-effectief	Factor kosten	Factor differen-tiatie	Uitkomst
Laag	20%	66%	1,0	25%	n.v.t.	25%	63%	0,5%
Midden	20%	66%	1,5	33%	n.v.t.	63%	100%	4,1%
Hoog	20%	66%	2,0	40%	n.v.t.	63%	100%	6,6%
Toelichting		Midden	Alleen medical reversals bekeken	Laag tot hoog		Laag tot midden	Midden tot hoog	

Tabel 9 De input uit Herrera-Perez is een waarde voor niet-effectief binnen niet-bewezen en kent vertaalslagen voor de overige factoren.

Schwartz: 1 tot 12%

Schwartz et al. schat dat stoppen met 26 veelgebruikte lage toegevoegde waarde behandelingen een significant deel van de uitgaven kan besparen (Schwartz & al., 2014). Het gaat hier om medisch specialistische zorg in het Amerikaanse Medicare (part A en part B) dat vooral voor ouderen relevant is. De besproken ingrepen werden gedaan bij 25% tot 42% van de patiënten.

Het doel van het onderzoek was om te onderzoeken of uit administratieve gegevens ('medical claims data') indicaties voor overgebruik van zorg bevatten. De auteurs selecteerden 26 ingrepen die concreet genoeg zijn, en waarover voldoende gegevens tussen ziekenhuis en Medicare worden uitgewisseld om een oordeel over gepast gebruik te geven. De 26 ingrepen zijn geselecteerd uit een longlist van verrichtingen die volgens vijf bronnen van lage waarden zijn (Choosing Wisely, US Preventive Services Task Force "D" aanbevelingen, "do not do" aanbevelingen van het National Institute of Clinical Excellence, het Canadese Agency for Drugs and Technologies in Health, en peer reviewed Australisch onderzoek).

Als input voor het model nemen we dat het aandeel niet-effectieve zorg 4,8% tot 32,4% is. Dat is een opscha-ling van de in de studie genoemde 0,6% tot 2,7% als aandeel van de totale zorgkosten. De reden hiervoor is dat we verwachten dat het totale aandeel niet-effectieve zorg een factor acht tot twaalf hoger is dan het onderzochte aandeel. Dit betreft een zeer ruwe aanname. Het programma ZE&GG ontving voor een inventarisatie van de top 10 handelingsagenda bijvoorbeeld 268 kennisvragen, wat een factor tien hoger is de in het onderzoek gebruikte 26 verrichtingen. Ook in andere overzichten van mogelijk en deels niet-effectieve behandelingen vinden we steeds enkele honderden. Merk op dat dit onderzoek een andere aanpak kent dan bij Badgery-Parker,

omdat Schwartz de uitgaven aan niet-effectieve zorg deelt door de totale zorgkosten, terwijl Badgery-Parker het deelt door de kosten van alleen de betreffende verrichtingen.

In het model vertalen we de input naar een schatting van 1,7% tot 11,5% voor Nederland (Tabel 10). Daarbij spelen nog de factor Nederland (50% voor vertaling vanuit VS) en opschaling van deelgebied, waarvoor we de factor van Beeldvorming en Diagnostiek gebruiken, omdat verreweg de meeste verrichtingen in Schwartz onder deze categorie vallen. De factor differentiatie is door de auteurs zelf al meegenomen in het lage en hoge scenario.

Scenario	Input	Factor Nederland	Factor opschaling	Factor niet-bewezen	Factor niet-effectief	Factor kosten	Factor differentiatie	Uitkomst
Laag	4,8%	50%	0,71	n.v.t.	n.v.t.	n.v.t.	n.v.t.	1,7%
Midden	18,6%	50%	0,71	n.v.t.	n.v.t.	n.v.t.	n.v.t.	6,6%
Hoog	32,4%	50%	0,71	n.v.t.	n.v.t.	n.v.t.	n.v.t.	11,5%
Toelichting		Laag	Beeldvorming en diagnostiek					

Tabel 10 De input uit Schwartz verlagen we met een factor 0,35 (=50%*0,71).

Clinical evidence: 4 tot 12%

In meerdere Nederlandse publicaties die ingaan op de opbrengst van zorgevaluatie en stringenter pakketbeheer, wordt verwezen naar het onderzoeksproject 'Clinical Evidence' van het British Medical Journal (BMJ). Uit dat onderzoek waarin 3.000 'common medical procedures' zijn ingedeeld naar bewijs van effectiviteit, wordt geconcludeerd dat voor de helft van de zorg de effectiviteit niet bekend is.

Deze conclusie hecht teveel waarde aan dat onderzoek. Het betreft geen formele peer reviewed artikel in BMJ, maar een korte update van het project om 'Clinical Evidence'⁴. Nederlandse rapporten gebruiken als bron vaak een Letter to BMJ uit 2007, getiteld "How much of orthodox medicine is evidence based" (Garrow, 2007) waarin de auteur J.S. Garrow verwijst naar het Clinical Evidence project. De originele bron is niet meer te vinden, maar wel op een archiefwebsite⁵ (BMJ, 2012). De volgende punten volgen uit dat (korte) webpost die namens het project geschreven lijkt te zijn en waarvan de auteur niet eenvoudig te achterhalen is.

- De schrijver(s?) geeft aan dat het subjectieve oordelen betreft, en dat een verrichting meerdere indicaties kan hebben, en voor sommige indicaties niet effectief en voor andere indicaties juist wel effectief kan zijn.
- De categorie 'niet bekende effectiviteit' wordt door de auteurs genuanceerd. Er vallen ook behandelingen onder waarvoor RCTs überhaupt niet geschikt zijn. Voor veel bekende

⁴ Het project is gestopt. BMJ is in plaats daarvan het project 'BMJ Best Practice' gestart. Alle artikelen die in het project "Clinical Evidence" zijn hier te vinden: "[BMJ Clinical Evidence](#)"[Journal] - PMC - NCBI ([nih.gov](#))

⁵ Dank aan Pauline de Heer (Zorginstituut) die ons hier op wees.

verrichtingen die in de praktijk goed blijken te werken, zal een medisch ethische toetsingscommissie geen toestemming geven om een RCT te doen.

- Onder het onderzoek vallen veel verrichtingen van alternatieve therapieën, zoals acupunctuur voor lage-rugpijn en innemen van echinacea (een kruid) bij verkoudheid. Deze zorg valt niet onder het basispakket van de zorgverzekeringswet.
- Een deel van de 3.000 onderzochte verrichtingen valt onder de eerstelijns zorg.
- De schrijvers benadrukken dat hun categorisering niet aangeeft hoe vaak de al dan niet bewezen effectieve verrichtingen in de praktijk gebruikt worden. Het betreft een aandeel van een telling van verrichtingen en is dus niet te gebruiken als aandeel voor zorguitgaven. De vaak geciteerde percentages zijn dus niet gewogen naar aantallen patiënten, noch naar uitgaven.
- De schrijvers benadrukken ook dat het onderzoek geen oordeel geeft over de effectiviteit voor individuele patiënten. In onze terminologie gaat het dus vaak om gedifferentieerde effectiviteit.

Bovenstaande punten worden nog versterkt door een discussie hierover in BMJ. Op de veel geciteerde brief van Garrow volgt een kennelijk minder gelezen brief in dezelfde uitgave van BMJ. Het betreft een reactie van D.I. Tovey, editorial director BMJ Knowledge⁶ (Tovey, 2007). Hij stelt: “The figures quoted by Garrow (previous letter) from BMJ Clinical Evidence are correct, but they provide an overly pessimistic picture of the state of evidence for orthodox medicine, since interventions in complementary and alternative medicine (CAM) and other non-orthodox treatments are included and are over-represented in the “unknown” category”. Overigens stelt hij ook “None the less, the general point is correct that many orthodox and complementary interventions are in common use despite uncertainty about their effectiveness.”

Met het rekenmodel vertalen we de studie naar 3,8% tot 11,4% niet effectieve zorg in Nederland (Tabel 11). Daarbij gebruiken we de factor 100% voor opschaling van bevindingen uit VK. De opschalingsfactor is relatief laag, omdat de studie voornamelijk om veelvoorkomende niet-complexe ingrepen lijkt te gaan.

Scenario	Input	Factor Nederland	Factor opschaling	Factor niet-bewezen	Factor niet-effectief	Factor kosten	Factor differentiatie	Uitkomst
Laag	40%	100%	0,71	n.v.t.	20%	100%	100%	5,7%
Midden	40%	100%	0,71	n.v.t.	30%	100%	100%	8,6%
Hoog	40%	100%	0,71	n.v.t.	40%	100%	100%	11,4%
Toelichting		VK	NC – veel voorkomend		Laag tot hoog			

Tabel 11 De input uit Clinical Evidence zegt alleen iets over de factor niet-bewezen en is dus afhankelijk van veel andere factoren.

⁶ Het lijkt erop dat hij verantwoordelijk was voor, dan wel meegewerkt heeft aan het Clinical Evidence project van BMJ.

Berwick: 3,5 tot 5%

Don Berwick en Andrew Hackbarth (Berwick & Hackbarth, 2012) stelden in 2012 dat de msz in de Verenigde Staten 20% kan worden bespaard door 'waste' terug te dringen. Een deel daarvan is vanwege niet-effectieve zorg.

We leiden af dat 7,0% tot 9,8% van de zorg in de VS niet-effectief is.

De studie gaat uit van zes categorieën 'waste', waarvan 'Overtreatment' het beste bij ons onderzoek naar niet-effectieve zorg past. De impact daarvan schatten de auteurs tussen 6,0% en 8,5% voor de VS. Mogelijk valt ook een deel van de categorie 'Failures of care delivery' onder onze definitie van niet-effectieve zorg. De opbrengst daarvan in de VS schatten de auteurs tussen 3,8% en 5,8%. Als we een kwart daarvan meenemen, kunnen we de schatting met 1,0% tot 1,4% verhogen. Mogelijke besparingen uit de overige categorieën rekenen we niet mee (dat zijn: failures of care coordination, administrative complexity, pricing failures, fraud and abuse).

Vertaald naar Nederland schatten we op basis van deze studie dat het aandeel niet-effectieve zorg 3,5% tot 4,9% is (Tabel 12).

Scenario	Input	Factor Nederland	Factor opschaaling	Factor niet-bewezen	Factor niet-effectief	Factor kosten	Factor differentiatie	Uitkomst
Laag	7,0%	50%	n.v.t.	n.v.t.	n.v.t.	n.v.t.	n.v.t.	3,5%
Midden	8,2%	50%	n.v.t.	n.v.t.	n.v.t.	n.v.t.	n.v.t.	4,2%
Hoog	9,8%	50%	n.v.t.	n.v.t.	n.v.t.	n.v.t.	n.v.t.	4,9%
Toelichting		Laag						

Tabel 12 De input uit Berwick verlagen we met een factor 0,5.

The Healthcare imperative: 4 tot 10%

Het Amerikaanse Institute of Medicine hield in 2010 een serie workshops over hoe de gezondheidszorg in de VS met minder uitgaven een beter resultaat kan halen (Institute of Medicine, 2010). Twee bijdrages aan de workshop gaan in op praktijkvariatie bij behandelingen die voor een deel van de patiënten niet-effectief zijn. Hieruit leiden we af dat 8% tot 20% van de zorg in de VS niet-effectief is:

- Op basis van onderzoek naar verschillen in uitgaven en uitkomsten in werkgebieden van ziekenhuizen in de VS voor AMI, heupfractuur en colorectale kanker ("Saving money and lives", Chandra et al, p86) concluderen de auteurs dat voor deze behandelingen 8% kosten bespaard kunnen worden, zonder dat de kwaliteit van zorg achteruit gaat (of zelfs verbetert). Ze suggereren dat dit percentage mogelijk op de totale zorg van toepassing is. Wij gaan ervan uit dat het aandeel niet-effectieve zorg in Nederland de helft van die in de VS is.
- Uit regionale benchmarks met het kwintiel of deciel met de minste uitgaven volgt 18 tot 20% niet-effectieve zorg ("Regional insights and US Health care savings", Fisher et al, p 95). Het ging om zorg in de zes laatste maanden van het leven van patiënten (tegen gestandaardiseerde prijzen en correctie voor demografische kenmerken, Medicare patiënten

>65 die niet waren ingeschreven bij een HMO). Volgens de auteurs zijn de uitgaven van een ziekenhuis aan patiënten jonger dan en ouder dan 65 gecorreleerd. Zij stellen dus dat het resultaat geldt voor alle leeftijden, en ook voor zorg niet aan het einde van het leven. Stel we gaan daar in mee, dan nog doen we een afslag voor de Nederlandse situatie die meer lijkt op een HMO dan op Medicare. Ook gemiddeld genomen, zijn er in Nederland veel minder verrichtingen per hoofd van de bevolking dan in de VS. Nemen we de helft van dit resultaat, dan komen we op 9 tot 10%.

- In een andere publicatie vertaalt het IOM zelf de uitkomsten van de workshop in \$210 miljard aan 'Unnecessary Services' in 2009 (Institute of Medicine, 2013). Dat is ongeveer 10% van de totale uitgaven aan zorg. Vertaald naar Nederland is dit 5%.

Hun bevindingen vertalen we naar de Nederlandse situatie als een potentie van grofweg 4% tot 10% minder uitgaven (Tabel 13).

Scenario	Input	Factor Nederland	Factor opschaaling	Factor niet-bewezen	Factor niet-effectief	Factor kosten	Factor differentiatie	Uitkomst
Laag	8%	50%	n.v.t.	n.v.t.	n.v.t.	n.v.t.	n.v.t.	4%
Midden	14%	50%	n.v.t.	n.v.t.	n.v.t.	n.v.t.	n.v.t.	7%
Hoog	20%	50%	n.v.t.	n.v.t.	n.v.t.	n.v.t.	n.v.t.	10%
Toelichting		Laag						

Tabel 13 De input uit The Healthcare Imperative verlagen we met een factor 0,5.

PwC: 5 tot 9%

In een onderzoek uit 2010 naar 'waste' in de Amerikaanse ziekenhuiszorg, schat PwC's Health Research Institute dat er ongeveer \$312 miljard van de \$2.200 miljard uitgaven aan curatieve zorg vallen onder 'clinical waste' is (PwC Health Research Institute, 2010).

Binnen de categorie 'clinical waste' zouden we de impact van twee subcategorieën geheel en een gedeeltelijk kunnen scharen onder niet-effectieve zorg. Vermijdbare uitgaven aan de meegetelde subcategorieën zijn ongeveer \$77 miljard tot \$ 129 miljard ofwel 3,5% tot 5,8% van de uitgaven.

- Subcategorieën met niet-effectieve zorg:
 - Unnecessary ER visits (\$14 mld),
 - Treatment variations (\$10 mld).
- De subcategorie 'defensive medicine' (\$210 miljard) bestaat uit overbodige behandelingen, vaak diagnostiek, die gedaan worden om claims voor bijvoorbeeld een gemiste diagnose te voorkomen. Het totale effect van 'defensive medicine' is onderwerp van discussie (zie bijvoorbeeld (Carrol, 2017)). Een deel van de niet-effectieve zorg valt hier waarschijnlijk onder. In Nederland speelt het probleem van defensive medicine minder en wordt bovendien relatief weinig beeldvormende diagnostiek gedaan. We nemen 25% tot 50%, ofwel \$53 tot \$105 miljard hiervan mee.

Naast 'clinical waste' noemde PwC 'behavioral waste' en 'operational waste', met respectievelijk ongeveer \$400 miljard en \$220 miljard aan uitgaven die volgens het onderzoek niet nodig zijn.

Omdat PwC clinical waste heeft gescheiden van andere soorten van waste, zijn de daadwerkelijke kosten van niet-effectieve behandelingen (die ook andere soorten waste met zich meedragen) hoger. PwC schat dat de helft van de uitgaven aan zorg in de VS 'waste' betreffen. We gaan ervan uit dat behavioral en operational waste in Nederland veel minder voorkomt (zie §2.2.1). Als er helemaal geen vergelijkbare 'waste' in Nederland zou zijn, zou de noemer halveren. We gaan uit van de helft tot een derde van de waste. De noemer is dan drie kwart ($1 - \frac{1}{2} * \frac{1}{2}$) tot twee derde ($1 - \frac{1}{2} * \frac{2}{3}$) lager. Dan schatten we dat dit onderzoek van PwC, vertaald naar de Nederlandse situatie dus $\frac{4}{3}$ tot $\frac{3}{2}$ maal zo hoog. Het aandeel niet-effectieve zorg ligt dan tussen 4,7% en 8,7% (uit $3,5 * \frac{4}{3}$ en $5,8 * \frac{3}{2}$) (Tabel 14).

Scenario	Input	Factor Nederland	Factor opscha- ling	Factor niet- bewezen	Factor niet- effectief	Factor kosten	Factor differen- tiatie	Uitkomst
Laag	3,5%	Inclusief	4/3	n.v.t.	n.v.t.	n.v.t.	n.v.t.	4,7%
Midden	4,7%	Inclusief	1,4	n.v.t.	n.v.t.	n.v.t.	n.v.t.	6,6%
Hoog	5,8%	Inclusief	3/2	n.v.t.	n.v.t.	n.v.t.	n.v.t.	8,7%
Toelichting		Aangepast						

Tabel 14 De input uit PwC kent een aangepaste vertaling naar Nederland en opscha-ling.

KPMG Plexus: 5 tot 7%

In 2010 onderzocht KPMG Plexus in opdracht van het Ministerie van VWS de budgettaire effecten van scherpere indicatiestelling (KPMG Plexus, 2010). Zij gaan ervan uit dat er in Nederland nauwelijks onderbehandeling is. Het grootste deel van praktijkvariatie betreft dan overbehandeling. Dat genereert onnodige zorgkosten.

De impact daarvan op het budgettaire kader medisch specialistische zorg (het rapport spreekt over 'macrokader') geschat op een potentiële besparing van 5% tot 7% (Tabel 15). Dit is gebaseerd op een onderzoek naar praktijkvariatie in Nederland voor behandelingen van vier aandoeningen: benigne prostaat hyperplasie (BPH), cataract, Hernia nucleii pulposi (HNP) en Ziekten van adenoid en tonsillen (ZAT). De schatting gaat uit van het maximaal terugdringen van de variatie richting gepaste referentiewaarden. Er is geen afslag gedaan voor implementatie. Wel zijn de referentiewaarden voor wel of niet indiceren na gesprekken met de wetenschappelijke verenigingen verlaagd tot realistisch haalbare niveaus.

Scenario	Input	Factor Nederland	Factor opschaling	Factor niet-bewezen	Factor niet-effectief	Factor kosten	Factor differentiatie	Uitkomst
Laag	5%	n.v.t.	n.v.t.	n.v.t.	n.v.t.	n.v.t.	n.v.t.	5%
Midden	6%	n.v.t.	n.v.t.	n.v.t.	n.v.t.	n.v.t.	n.v.t.	6%
Hoog	7%	n.v.t.	n.v.t.	n.v.t.	n.v.t.	n.v.t.	n.v.t.	7%

Tabel 15 De input uit KPMG Plexus is direct over te nemen.

2.2.3 Gevoeligheidsanalyse: schatting meest gevoelig voor vertaling onderzoeken naar Nederland en opschalingen uit deelgebieden – bandbreedte 1 tot 12%

Voor het merendeel van de geanalyseerde onderzoeken moesten we meerdere onzekere aannames doen om de resultaten te vertalen naar de Nederlandse medisch specialistische zorg. In Tabel 16 is per onderzoek aangegeven welke factoren onzeker zijn en invloed hebben op de uitkomsten. Naast de factoren uit het rekenmodel is daar nog een kolom aan toegevoegd die aangeeft of de input-percentages uit het onderzoek gebaseerd zijn op een telling van verrichtingen die niet gewogen is naar patiëntenaantallen en kosten.

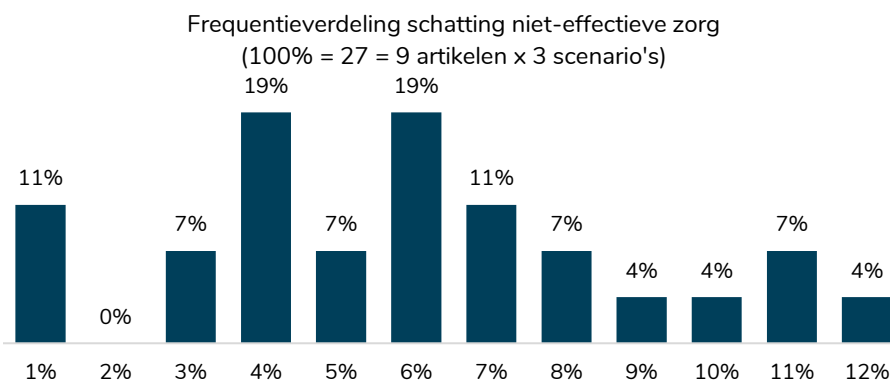
Onderzoek	Factor Nederland	Factor opschaling	Factor niet-bewezen	Factor niet-effectief	Factor kosten	Factor differentiatie	Onzekerheid door kale telling
Badgery-Parker	v	v			v		
Prasad	v	v	v		v	v	v
Herrera-Perez	v	v	v		v	v	v
Schwartz	v	v					
Clinical Evidence	v	v		v			v
Berwick	v						
The Healthcare Imperative	v						
PwC	v						
KPMG Plexus		v					

Tabel 16 Overzicht onzekerheden per onderzoek: de meest voorkomende onzekerheden zijn de factor Nederland en de factor opschaling.

Het voordeel van het gebruik van meerdere verschillende bronnen is dat niet elke bron van alle aannames afhankelijk is. De interpretatie van Prasad, Herrera-Perez en Clinical Evidence is bijvoorbeeld afhankelijk van de vertaling van aandelen niet-effectieve zorg in een telling van verrichtingen of publicaties naar een kostenaandeel. Dit kent een grote onzekerheid, die we in dit rapport helaas niet kunnen kwantificeren. De andere studies onderzoeken juist expliciet de kosten van niet-effectieve behandelingen, door ofwel de kosten van niet-effectieve behandelingen stuk voor stuk op te tellen (Badgery-Parker, Schwartz) ofwel door te kijken naar praktijkvariatie in de

kosten voor behandelingen waarvan bekend is dat zij voor een deel van de populatie niet-effectief zijn (gebruikt voor de vier laatstgenoemde onderzoeken).

De grootste onzekerheid komt door de vertaling van internationaal onderzoek naar Nederland en de opschaling van onderzoek in deelgebieden naar de totale zorg. Als in deze aannames een systematische onder- of overschatting zit, heeft dit direct effect op de gepresenteerde resultaten. Een deel van deze onzekerheid wordt wel meegenomen in de verschillende scenario's en komt daarom deels terug in de totale bandbreedte van 1% tot 12% (Figuur 9).



Figuur 9. Bandbreedte van 1 tot 12% voor schatting van aandeel niet-effectieve zorg met gemiddelde (en mediaan) van 6,2%.

Deze conclusie sluit aan bij het beeld uit interviews: er heeft simpelweg nog geen grootschalige studie naar het aandeel niet-effectieve msz in Nederland plaatsgevonden. Tot die tijd zijn we dus afhankelijk van (onzekere) vertalingen van studies uit het buitenland en/of opschalingen van resultaten in deelgebieden.

2.3 Niet-effectieve zorg groeit jaarlijks met ongeveer 0,3%-punt

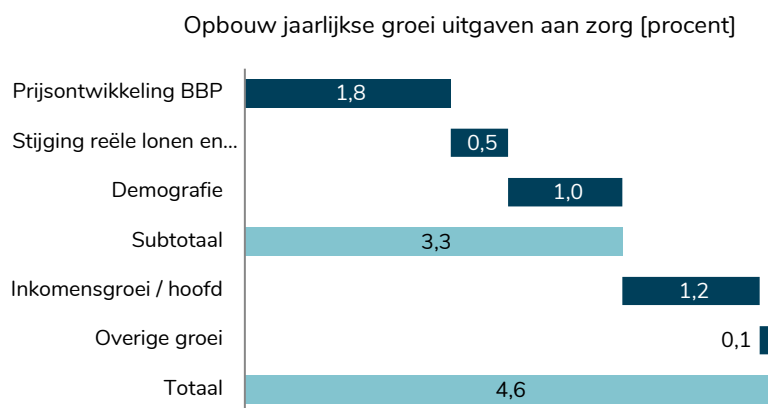
We schatten de jaarlijkse uitgavengroei door technologie en indicatieverruiming op ongeveer 1,5%. Daarvan zou ongeveer 20% niet-effectief kunnen zijn. Samen dus 0,3% van de €24,3 miljard aan uitgaven aan medisch specialistische zorg.

2.3.1 Jaarlijkse instroom van 1,5% aan uitgaven aan 'nieuwe' zorg

Jaarlijks stroomt er naar schatting circa 1,5% van uitgaven aan medisch specialistische zorg als 'nieuwe zorg' het pakket in (Figuur 10). Hieronder verstaan we een combinatie van indicatieverruiming en nieuwe technologie. Dit is op basis van een synthese van meerdere bronnen:

- Als we uitgaan van het deel van de jaarlijkse groei aan zorguitgaven dat is toe te schrijven aan de inkomstengroei per hoofd van de bevolking en overige groei komen we uit op totaal 1,3% (Centraal Planbureau, 2019). De algemene verwachting is dat dit komt door instroom van (dure) technologie en indicatieverruiming. Het betreft dan overigens wel het saldo van die instroom met productiviteitsverbeteringen, dus is de 1,3% een ondergrens. De overige factoren prijsontwikkeling, stijging van reële lonen en prijzen en demografie, verklaren de groei met de huidige manier van werken.

- Dit komt redelijk overeen met cijfers van de (Europese Commissie, 2010), die uitgaan van 2,0% jaarlijkse groei door technologische ontwikkelingen. Dat beschouwen wij als een bovengrens, mede omdat dit moeilijk af te bakenen is, zoals de auteurs zelf ook benadrukken.
- (Sorensen, Drummond, & Bhuiyan Khan, 2013) schatten op basis van een review van 86 studies dat 50% (25%-75%) van de reële groei door technologie komt. Dit lijkt ook consistent met de bespreking van Chernew & Newhouse in het Handbook of Health Economics (Chernew & Newhouse, 2012).



Figuur 10 Jaarlijkse groei zorguitgaven bestaat voor 1,3%-punt uit inkomensgroei per hoofd en overige groei. Bron: decompositie van CPB op basis van ontwikkeling vanaf 1990, zie CPB Middellangetermijnverkenning Zorg 2022-2025.

2.3.2 Ongeveer een vijfde van de nieuwe instroom is zonder actief pakketbeheer niet-effectief

Helaas is maar weinig bekend over de werkelijke effectiviteit van zorg die het pakket instroomt. Het meeste onderzoek daartoe wordt gedaan voor geneesmiddelen. We schatten dat 20% van de nieuwe zorg (qua omzet) zonder actief pakketbeheer niet-effectief is. Uit internationale studies volgt namelijk dat voor circa 80% van de nieuwe instroom geconcludeerd kan worden dat de zorg effectief is:

- Tussen 2000 en 2010 liet het Engelse National Institute for Clinical Excellence (NICE) volgens Garner et al. 83% van de onderzochte technologieën toe in de National Health Service (NHS) (Garner & Littlejohns, 2011).
- Volgens Prasad et al. (Prasad, et al., 2013) bleek dat 77% van de RCT's die nieuwe verrichtingen onderzochten, concludeerden dat die verrichtingen effectief zijn.

In de volgende hoofdstukken gaan we verder in op het onderzoek dat nodig is om deze zorg te identificeren (zie in het bijzonder §4.2).

Gezamenlijk concluderen we dat jaarlijks ongeveer 0,3% van de uitgaven aan msz ($20\% * 1,5\%$) niet-effectieve zorg het pakket instroomt zonder actief pakketbeheer.

2.4 Geïnterviewden benadrukken belang van effectief pakketbeheer, maar zijn terughoudend over besparingspotentieel

Uit de interviews en websessies, waarin bovenstaande aanpak en resultaten besproken zijn, concluderen we:

- De schatting blijft onvermijdelijk afhankelijk van veel ruwe en onzekere aannames.
- De meerkosten van niet-effectieve zorg in het systeem leiden niet direct tot besparingspotentieel.
- Het doel van effectief pakketbeheer zou niet in de eerste plaats kostenbesparing, maar het verbeteren van de kwaliteit van zorg moeten zijn.

Geïnterviewden benadrukken dat het met de huidige kennis onmogelijk is een precieze schatting van het aandeel niet-effectieve msz te doen. Zij herkennen wel de orde van grootte en de algemene richting van de aannames die we hebben moeten doen. Grootschalig en betrouwbaar onderzoek naar het aandeel niet-effectieve zorg heeft nog niet plaatsgevonden. De huidige kennis is voornamelijk gegenereerd door de individuele betrokkenheid en inzet van medisch specialisten, waardoor de bekende voorbeelden geen representatief beeld van de totale msz geven.

Ook geven geïnterviewden aan dat meerkosten niet direct tot besparingspotentieel leiden. De uitgaven aan ziekenhuiskosten kennen veel vaste kosten, die niet op korte termijn afgebouwd kunnen worden. Mogelijk kan de kostenbasis zich wel in termijn van vijf of meerdere jaren aanpassen, bijvoorbeeld door concentratie van zorg of faciliteiten zoals laboratoria.

Ze geven aan dat het voorkomen van niet-effectieve behandelingen ruimte creëert voor een toename van andere, effectieve zorg binnen eenzelfde macrokader. Met de juiste financiële prikkels kan daarom ingezet worden op een vervanging van niet-effectieve door effectieve zorg binnen het kader.

Verder benadrukken geïnterviewden dat effectief pakketbeheer in de eerste plaats als doel zou moeten hebben de kwaliteit van zorg te verbeteren. Niet-effectieve zorg kan schadelijk zijn voor de patiënt, terwijl voor specifieke patiënten een andere vorm van zorg mogelijk wel effectief blijkt. In het verlengde hiervan geven geïnterviewden aan dat het doelmatiger organiseren van het zorgaanbod vermoedelijk tot meer kostenbesparing kan leiden dan effectiviteitsonderzoek.

3 Circa 80% van de niet-effectieve zorg is te vermijden

We verwachten dat circa 80% van de niet-effectieve zorg te vermijden is. Zonder wetswijzigingen gaan we ervan uit dat het met forse inspanning lukt om te stoppen met 60% van de niet-effectieve zorg (§3.1). Mét wetswijzigingen denken we dat de ambitie van 80% die de HLA-partijen voor het programma Zorgevaluatie en Gepast Gebruik (ZE&GG) hebben uitgesproken, haalbaar is (§3.2). Voor het vermijden van 100% van de niet-effectieve zorg zouden in de hele Nederlandse zorg ideale omstandigheden gecreëerd moeten worden (§3.3).

In Hoofdstuk 2 schatten we dat circa 6% van de uitgaven van zorg vermeden kan worden door geen niet-effectieve zorg meer te leveren. Het blijkt echter in de praktijk een forse uitdaging om de indicatiestelling aan te scherpen en geen niet-effectieve zorg meer te leveren. Deimplementatie van niet-effectieve zorg blijkt lastig te organiseren, terwijl het cruciaal is om onnodige zorg en daarmee kans op onnodige schade aan patiënten te voorkomen (Norton & Chambers, 2020). Alhoewel het onderwerp steeds meer aandacht krijgt, wordt er nog vrij weinig onderzoek naar deimplementatie gedaan (Norton, Kennedy, & Chambers, 2017). Zo wordt bijvoorbeeld opgeroepen om systematisch onderzoek te doen naar de impact van de Choosing Wisely campagnes (Bhatia, W, & Shortt S, 2015). Bevindingen van door ZonMw gesubsidieerd doelmatigheidsonderzoek leiden ook in Nederland lang niet altijd tot veranderingen in de praktijk, of tot opname in richtlijnen (SiRM, 2018). Ook de ervaringen met de introductie van nieuwe kapitaalintensievere technologieën zoals robotchirurgie en protonentherapie, laten zien dat het beheersen van dergelijke processen erg moeilijk is (CPB, 2017).

Om te bepalen welk aandeel van de levering van de 6% niet-effectieve zorg ook daadwerkelijk voorkomen kan worden, werken we met drie scenario's. Deze scenario's zijn in principe onafhankelijk van de scenario's voor het tempo van implementatie uit Hoofdstuk 4.

- Een laag scenario van 60%.
- Een midden scenario van 80%.
- Een bovengrens van 100%.

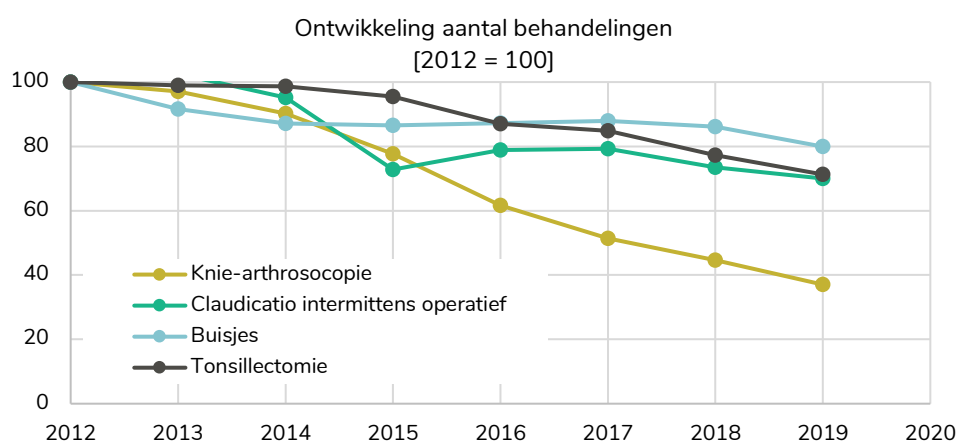
Bij deze scenario's horen stevige maatregelen omtrent wet- en regelgeving en investeringen in zorgevaluatie. We gaan in Hoofdstuk 4 in op het tempo waarmee dat doel gehaald kan worden en op de bijbehorende uitgaven.

3.1 Met forse inspanning – maar zonder wetswijzigingen – lukt het wellicht 60% van de niet-effectieve zorg te vermijden

We gaan in het lage scenario uit van 60%. Deze 60% is geïnspireerd op de daling die in circa een decennium is gehaald van knie-artroscopieën die niet altijd effectief bleken, en een optimistische interpretatie van de daling in tonsillectomieën. Andere chirurgische ingrepen lieten minder grote dalingen zien. Er is dus ook om het doel van 60% te behalen, een forse inspanning nodig. Denk aan voldoende middelen om het benodigde onderzoek naar effectiviteit te agenderen, uit te voeren en de inzichten in het veld door te voeren. De BMH gingen uit van totaal €120 miljoen per jaar. Wij gaan hier in Hoofdstuk 4 verder op in.

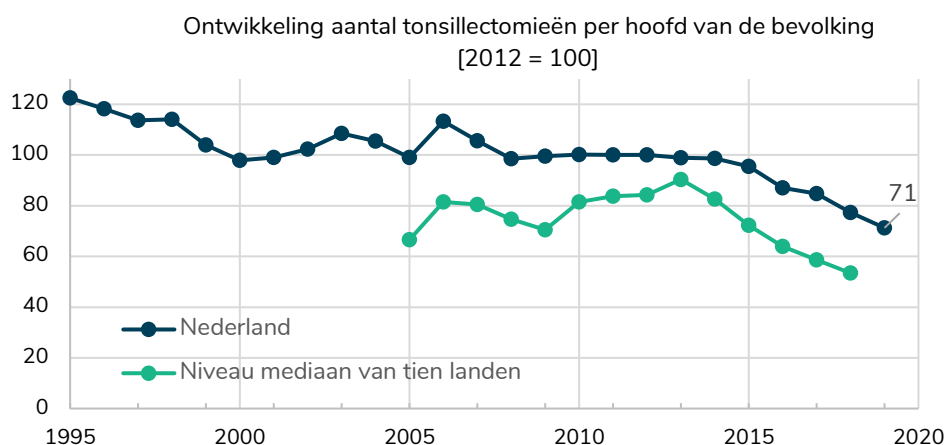
We onderzochten enkele chirurgische ingrepen waarvan al voor 2010 bekend was dat ze niet altijd effectief zijn. Het aantal ingrepen onderging een daling van 20% tot 60% tussen 2012 en 2019 (Figuur 11). Voor de knie-arthroscopie was de daling met 60% het sterkst.

Mogelijk is hier een groter deel van de te vermijden zorg mee gerealiseerd omdat het uiteindelijke optimale niveau niet op nul zou liggen. In geen van de grafieken is vooralsnog een asymptoot naar het effectieve niveau te herkennen (Figuur 11).



Figuur 11. Behandelingen waarvan bekend is dat ze niet voor alle patiënten effectief zijn, laten nog geen asymptoot zien na ongeveer vijf jaar daling (bron: SiRM analyse obv OpenDis gegevens).

Voor tonsillectomie is ook informatie over eerdere jaren bekend. We zien dan dat het decennia kan duren voordat het aantal behandelingen daalt. Uit internationale vergelijking blijkt bovendien dat het effectieve niveau waarschijnlijk nog niet behaald is. Sinds 2005 daalt overal het aantal tonsillectomieën per hoofd van de bevolking. De mediaan van tien West-Europese landen met enigszins vergelijkbaar klimaat, lag in 2018 op 70% van het Nederlandse aantal (Figuur 12). Als dat het doelniveau zou zijn, werd daarvan ten opzichte van het plateau tussen 2008 en 2014 in 2019 60% bereikt ($((100-71)/(100-50))$).



Figuur 12 Aantal tonsillectomieën is in Nederland nog niet op laagste niveau. Bron: SiRM analyse obv gegevens van de OESO over aantal ingrepen per hoofd van de bevolking over België, Denemarken, Duitsland, Finland, Ierland, Luxemburg, Noorwegen, Verenigd Koninkrijk, Zweden.

3.2 Mét wetswijziging lijkt ambitie ZE&GG om 80% van niet-effectieve zorg te vermijden haalbaar

HLA-partijen hebben voor het programma ZE&GG de ambitie uitgesproken om 80% van de niet-effectieve zorg te vermijden (Repping, 2019). Wij denken dat deze ambitie haalbaar is, maar hiervoor is wel een wetswijziging nodig.

Vrijwel alle geïnterviewden benadrukten dat een effectief basispakket alleen te realiseren is als alle partijen, en vooral de medisch specialisten actief bijdragen. Zij geven aan dat daarbij zowel een intensivering van de inspanningen van de overheid, als van de veldpartijen benodigd is. Dit werd ook in de werksessies bevestigd.

Omdat de rol van veldpartijen zo belangrijk is voor de haalbaarheid van het afschalen van niet-effectieve zorg, raden we aan om voort te bouwen op het programma Zorgevaluatie en Gepast Gebruik (ZE&GG). Hier gaan we in het volgende hoofdstuk verder op in.

We denken dat de ambitie van ZE&GG om op termijn 80% van de niet-effectieve zorg te vermijden, zeer uitdagend, maar wellicht haalbaar is. Daarvoor moeten wel aanpassingen aan wet- en regelgeving gedaan worden. Het gaat daarbij vooral om aanpassing van de standpunten die het Zorginstituut in kan nemen:

- Het Zorginstituut zou naast het standpunt 'effectief' of 'niet-effectief' een standpunt 'nader onderzoek nodig' in moeten kunnen nemen. Een dergelijk standpunt betekent vervolgens dat die zorg alleen in een onderzoeksetting geleverd mag worden. Na een vooraf bepaalde periode, die afhangt van de patiëntenpopulatie en de behandeling, wordt op basis van het onderzoek besloten of de zorg als effectief kan worden beschouwd. De genoemde maatregelen zijn voor dit rapport niet getoetst op juridische haalbaarheid.
- Het Zorginstituut zou, op termijn en waar relevant, in standpunten over bestaande zorg meer handvatten moeten krijgen om grofmazig aan te geven voor welk aandeel van een brede patiëntenpopulatie verwacht wordt dat de zorg effectief is. Op die manier kunnen standpunten sneller tot stand komen dan wanneer op detailniveau bepaald moet worden

voor welke subpopulaties de zorg precies (niet-)effectief is. Zorgverzekeraars kunnen dat in hun contractering gebruiken. Hierbij wordt rekening gehouden met verschillen tussen aanbieders. Zo kan bijvoorbeeld bij een gespecialiseerde zorgaanbieder, het aandeel van een bepaalde behandeling juist hoger zijn, en bij een andere aanbieder lager of nul.

Overigens houdt deze uitbreiding van het repertoire aan standpunten tevens een sterke betrokkenheid van het veld in. De standpunten van het Zorginstituut zijn het formele sluitstuk van een traject dat door en voor het zorgveld wordt doorlopen. Het programma ZE&GG is daar goed voor geëquipeerd.

Naast de aanpassing van de mogelijke standpunten zijn nog meer maatregelen nodig op het gebied van management en implementatie om het doel te halen. We gaan daar bij het bespreken van het tempo van de maatregel in Hoofdstuk 4 verder op in.

3.3 Voor het vermijden van 100% zijn in de hele Nederlandse zorg ideale omstandigheden nodig

In het hoge scenario zouden alle resultaten van de evaluaties er in de praktijk toe leiden dat alle zorg effectief is. Dit is in theorie mogelijk, omdat we in de schatting in Hoofdstuk 2 rekening hebben gehouden met gedifferentieerde effectiviteit en dus alleen dat deel van de zorgverlening pogen mee te nemen dat daadwerkelijk niet-effectief is.

De praktijk is weerbarstiger. Om deze bovengrens te halen moet aan de voorwaarden voor het laag (§3.1) en midden scenario (§3.2) worden voldaan en daarnaast nog de 'ideale' zorgwereld zijn ontstaan. De kans daarop is klein gegeven bijvoorbeeld de volumeprikkels voor ziekenhuizen en medisch specialisten en onvolledige gelijkgerichtheid in de gehele zorgketen (SiRM, 2019).

Overigens zijn er wel eerste signalen dat het Beatrixziekenhuis (Gorinchem) en Bernhoven (Uden), met andere contracten met de zorgverzekeraar en weghalen van de volumeprikkel voor medisch specialisten, minder zorg produceren (CPB, 2020). De gerapporteerde daling van het behandelvolume is mogelijk deels veroorzaakt door minder niet-effectieve zorg te doen, maar een groot deel betreft waarschijnlijk ook efficiencyverbetering, prijsaanpassingen en substitutie van zorg.

4 Met voldoende investering kan merendeel in 15 jaar geïmplementeerd worden

We schatten in dat met voldoende investering het merendeel van de circa 80% niet-effectieve zorg in 15 jaar afgebouwd kan worden. Het vergt circa 15 jaar om vanuit aanpassingen in wet- en regelgeving en intensivering van inspanningen niet-effectieve zorg grotendeels af te bouwen (§4.1). Dit vergt een investering die tijdelijk oploopt tot circa €100 miljoen per jaar, en structureel circa €60 miljoen per jaar bedraagt (§4.2).

In dit hoofdstuk gaan we in op het tempo waarmee het afbouwen van niet-effectieve zorg geïmplementeerd kan worden en de extra uitgaven die daarvoor nodig zijn.

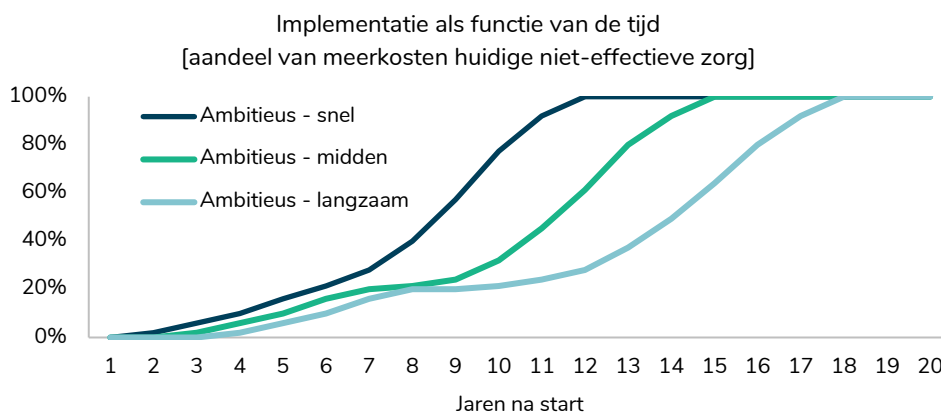
4.1 Het vergt circa 15 jaar om vanuit wetswijziging en intensivering van inspanningen niet-effectieve zorg grotendeels af te bouwen

Veldpartijen hebben, in het kader van afspraken die ze hierover maakten in het HLA, de afgelopen anderhalf jaar geïnvesteerd in het samenbrengen en intensiveren van inspanningen voor effectieve zorg. Ze hebben dit gedaan aan de hand van de zogenaamde cirkel van ZE&GG (Figuur 13). De stappen in de cirkel gebruiken we om het tempo van het vermijden van niet-effectieve zorg in kaart te brengen. We laten daar nog een opstartfase aan vooraf gaan in verband met de benodigde aanpassing van wet- en regelgeving (§3.2).



Figuur 13 Cirkel met de drie stappen van Zorgevaluatie en Gepast Gebruik. Goede systeeminbedding is hierbij een randvoorwaarde.

Met juiste wet- en regelgeving en intensivering van agenderen en evalueren, schatten we dat de het maximale potentieel aan vermeden niet-effectieve zorg in circa vijftien jaar bereikt wordt (Figuur 14).



Figuur 14 De implementatiegraad van het vermijden van niet-effectieve zorg stijgt van circa 10% na vijf jaar tot maximaal na vijftien jaar.

In de figuur houden we rekening met verschillende scenario's voor het tempo in de fases van de cirkel van ZE&GG (Tabel 17). We houden bovendien rekening met een opstartperiode van twee jaar voor wetswijzigingen en organisatorische aanpassingen voor het agenderen van de nog onbekende zorg. In onderstaande paragrafen lichten we per fase toe wat hiervoor nodig is.

Scenario	Opstart	Agenderen	Evalueren	Implementeren	Totaal
Snel	2 jaar	1 jaar	4 jaar	5 jaar	12 jaar
Midden	2 jaar	2 jaar	5 jaar	6 jaar	15 jaar
Langzaam	2 jaar	3 jaar	6 jaar	7 jaar	18 jaar

Tabel 17 Scenario's en bijbehorende doorlooptijden om niet-effectieve zorg af te schalen.

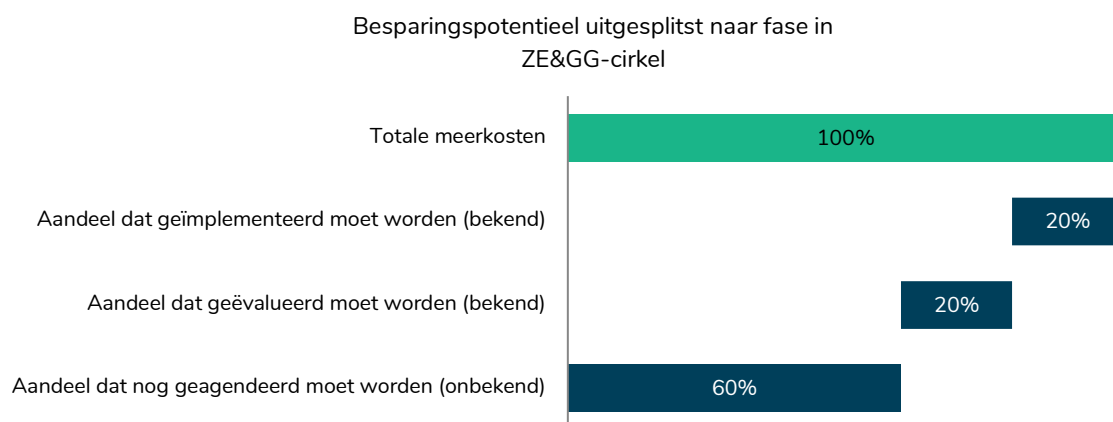
4.1.1 Opstartfase van twee jaar voor uitbreiding rol Zorginstituut, waarin lopende evaluaties en implementaties wel voortgang boeken

Uit de interviews maken we op dat er ongeveer twee jaar nodig is om het Zorginstituut volledig in zijn nieuwe rol te laten functioneren. De aanpassing van de wet om ook een standpunt 'onderzoek nodig' in te nemen duurt twee tot drie jaar. In de tussentijd kan de interne organisatie met extra financiering uitgerust worden om het pakket door te lichten op zorg met onbekende effectiviteit en hoge budgetimpact.

Tijdens de opstartfase kan al wel gestart worden met het (verder) afbouwen van bekende niet-effectieve zorg en de evaluatie van onderwerpen op de huidige kennisagenda's. We schatten dat van ~40% van de niet-effectieve zorg direct in de implementatie- of evaluatiefase geplaatst kan worden (Figuur 15):

- Uit Schwartz leiden we af dat de meerkosten van 26 veelvoorkomende niet-effectieve behandelingen 0,3% tot 1,4% van de uitgaven betreffen. Dit vertaalt zich naar 5% tot 23% van de totale geschatte meerkosten. Deze kunnen nu al geïmplementeerd worden.
- Op dit moment lopen er circa 43 evaluatieonderzoeken bij ZE&GG die niet afhankelijk zijn van bovenstaande wetswijzigingen. Het programma bestaat nog maar anderhalf jaar, dus het is de verwachting dat dit aantal de komende jaren zal toenemen. Vergeleken met de (ruw) opgeschaalde resultaten uit het rekenmodel op basis van Schwartz (§2.2.2) suggereert dit ook

een potentiële besparing van circa 1% van de ziekenhuisuitgaven, oftewel een kleine 20% van de totale meerkosten die zonder aanvullende agendering in de evaluatiefase komen.



Figuur 15 Circa 40% van het besparingspotentieel kan vermoedelijk zonder aanvullende agendering direct geëvalueerd of geïmplementeerd worden.

4.1.2 Termijn voor agenderen twee jaar door risicogericht onderwerpen voor zorgevaluatie te agenderen

Het is (op korte termijn) niet haalbaar om alle zorg waarvoor onzekerheid bestaat over de effectiviteit te onderzoeken. Er dient dus geprioriteerd te worden. Een manier van prioriteren is op basis van de budgetimpact die de zorg heeft.

We bevelen aan om het Zorginstituut te vragen om risicogericht onderwerpen voor zorgevaluatie te agenderen. Dit betekent dat het Zorginstituut gevraagd wordt om (samen met WVen) systematisch diagnostiek, behandeling en follow-up van patiënten met hoge budgetimpact in kaart te brengen, zowel voor bestaande als voor nieuwe zorg. Voor nieuwe zorg vergt dit het inrichten van een proces van horizon scanning, zoals dit al bestaat voor dure geneesmiddelen. Vervolgens gaat het Zorginstituut na of er voor deze zorg al voldoende wetenschappelijk bewijs bestaat over de effectiviteit ervan. Als voor de zorg wetenschappelijke onderbouwing ontbreekt, luidt het standpunt dat de zorg nog niet bewezen is en dat deze onderzocht dient te worden.

Op deze manier verwachten we dat het agenderen van onderzoek naar de zorg met onbekende effectiviteit en hoge budgetimpact in het huidige pakket in circa twee jaar uitgevoerd kan worden. Voor een snel scenario houden we rekening met één jaar en voor een langzaam scenario met drie jaar. In §4.2 gaan we op de bijkomende kosten in.

Een kanttekening bij bovenstaande focus op agendering van onderzoek naar het huidige pakket is dat veel zorguitbreidingen uit indicatieverruiming van bestaande zorg bestaan. Bestaande behandelingen zijn dus mogelijk ook in de toekomst onderdeel van de agenda.

4.1.3 Termijn voor evaluatie vijf jaar door versterken onderzoeksinfrastructuur

We gaan in onze berekening uit van een gemiddelde evaluatieperiode per onderzoek van vijf jaar. Vanaf het opzetten van de studie tot het publiceren van de resultaten lijkt vijf jaar ambitieus, maar dit is naar verwachting haalbaar door de beschreven wetswijziging (§3.2). Daardoor worden

immers alle aangewezen behandelingen in onderzoeksverband uitgevoerd, wat een van de belangrijkste knelpunten op dit moment – het includeren van voldoende patiënten in de evaluatieonderzoeken – op kan lossen. Deze termijn is tevens in lijn met de aannames uit Zorgkeuzes in Kaart en sluit ook aan bij de inschatting van geïnterviewden. Het betreft wel een gemiddelde - dit verschilt dus per onderzoek. In het meest optimistische scenario (snel) is de doorlooptijd mogelijk te versnellen naar vier jaar gemiddeld, terwijl we voor een langzamer scenario uitgaan van zes jaar.

Om de in Tabel 17 geschetste doorlooptijd te halen, is het daarnaast nodig dat er meer kwalitatief goede onderzoeken tegelijkertijd worden uitgevoerd dan op dit moment het geval is. Daartoe is verdere versterking van de onderzoeksinfrastructuur een vereiste. Denk hierbij aan meer samenwerking tussen trialbureaus om ervoor te zorgen dat niet voor ieder afzonderlijk onderzoeksproject een eigen kleine en tijdelijke structuur wordt opgezet, die na afloop van de zorgevaluatie weer wordt afgebroken. Deze samenwerking moet ook de kwaliteit van de onderzoeksvoorstellen verbeteren. Op dit moment is bijvoorbeeld bij een deel van de ingediende onderzoeksvoorstellen bij ZonMw de kwaliteit onvoldoende zoals bleek uit de geïdentificeerde verbeterpunten in een evaluatie van de regeling veelbelovende zorg. Het Zorginstituut kan hier ook een rol bij spelen. Daarnaast is het financieren van landelijke aandoeningsregisters van belang, waardoor relevante informatie voor het agenderen en uitvoeren van onderzoek sneller toegankelijk is. In §4.2 gaan we op de bijkomende kosten in.

Met deze aannames schatten we dat het benodigde onderzoek om het huidige pakket te evalueren in de komende tien jaar uitgevoerd kan worden, met een piek tussen jaar drie en acht. Daarna wordt een steady state bereikt waarin alleen de nieuwe instroom jaarlijks geëvalueerd moet worden. Op het hoogtepunt, circa vijf jaar na de start, is 80% van het benodigde onderzoek om het huidige pakket te evalueren gaande (Figuur 16).



Figuur 16 Om huidige pakket te evalueren is een sterke intensivering nodig, met een piek tussen jaar drie en jaar acht. Vanaf jaar tien is een steady state voor alleen nieuwe instroom bereikt. Figuur op basis van scenario midden.

4.1.4 Termijn voor implementatie zes jaar door inzet spiegelinformatie en vlotte richtlijnherziening

Het verminderen van gebruik van niet-effectieve zorg wordt op dit moment als groot knelpunt gezien (Hoofdstuk 3). In deze paragraaf gaan we in op de verhoging van het tempo van afschalen

van niet-effectieve zorg. We zien op dit moment drie redenen waarom het tempo in de toekomst omhoog kan:

- Het programma ZE&GG krijgt steeds meer aandacht.
- De FMS herzielt richtlijnen sneller.
- Meer spiegelinformatie over de implementatiegraad van het afbouwen van niet-effectieve zorg.

De eerste twee punten hebben zonder aanvullende maatregelen naar verwachting een positief effect op de implementatie. Veldpartijen hebben de afgelopen anderhalf jaar veel geïnvesteerd in het programma ZE&GG. Dit wordt de komende jaren verder uitgebouwd. ZE&GG werkt op dit moment bijvoorbeeld al aan implementatietoolkits met onder andere impactanalyses en spiegelinformatie.

De FMS werkt sinds kort met permanente commissies voor richtlijnherziening. Daardoor worden richtlijnen jaarlijks, modulair, aangepast. Voorheen kon het tot acht jaar duren voordat nieuwe inzichten vertaald werden in richtlijnen.

Het belangrijkste aandachtspunt voor de komende jaren is het derde punt – het verstrekken van spiegelinformatie. Het past binnen de ambitie van programma ZE&GG om deze spiegelinformatie voor alle partijen toegankelijk te maken. Met name in combinatie met de in §3.2 beschreven mogelijkheid voor het Zorginstituut om voor bepaalde handelingen gericht aan te geven voor welk percentage van de patiëntenpopulatie de behandeling effectief is, kan goede spiegelinformatie veel betekenen in de implementatiefase.

Zorgverzekeraars kunnen spiegelinformatie – in combinatie met een standpunt van het Zorginstituut – gebruiken in de contractering. Het valt onder de verantwoordelijkheid van zorgverzekeraars om gepast gebruik te stimuleren. Dit is retrospectief echter moeilijk op casusniveau te doen, omdat het meestal om gedifferentieerde effectiviteit gaat. Het oordeel hangt dan immers af van de situatie van de specifieke patiënt en is daarom moeilijk te geven. In de contractering kunnen echter wel op groepsniveau afspraken gemaakt worden over hoe vaak een bepaalde behandeling plaats kan vinden. Dit is naar verwachting een effectiever en efficiënter middel om gepast gebruik te stimuleren. Het verwachte aandeel patiënten dat een bepaalde behandeling krijgt, hoeft overigens niet voor alle zorgaanbieders gelijk te zijn. Uiteraard moet bij de contractering rekening gehouden worden met eventuele specialisatie. De NZa speelt een belangrijke rol in het stimuleren van zorgverzekeraars om deze rol te pakken.

Ook binnen ziekenhuizen dient gepast gebruik een structureel onderdeel van het gesprek te zijn tussen de Raad van Bestuur en medisch specialisten en tussen medisch specialisten onderling. De IGJ speelt een belangrijke rol in het vermijden van niet-effectieve zorg. Die is namelijk in de eerste plaats onnodig schadelijk voor de patiënt. Ook voor de IGJ is het vaak moeilijk om de effectiviteit van een individuele behandeling te beoordelen. Desondanks kan zij wel actief toezien op procesindicatoren voor het evalueren van gepast gebruik binnen ziekenhuizen.

Uitkomstindicatoren kunnen de IGJ extra handvatten bieden de effectiviteit van behandelingen te beoordelen.

Voor de implementatie gaan we met deze aanpassingen uit van een termijn van implementeren van vijf tot zeven jaar, bestaande uit één tot drie jaar voor de implementatie in richtlijnen en een termijn van vier jaar voor stapsgewijze implementatie in de praktijk. Deze vier jaar geeft zorgverzekeraars en ziekenhuisbestuurders naar verwachting voldoende jaren aan spiegelinformatie om de gewenste afbouw van niet-effectieve zorg te implementeren.

4.2 Benodigde extra uitgaven zijn tijdelijk circa €100 miljoen en structureel circa €60 miljoen per jaar

Voor de benodigde extra uitgaven onderscheiden we vaste en variabele kosten:

- De vaste kosten bestaan uit
 - extra uitgaven aan het Zorginstituut en
 - investeringen in de structurele onderzoeksinfrastructuur.
- De variabele kosten bestaan uit de kosten per evaluatieonderzoek.

In lijn met het de aannames van het CPB in Zorgkeuzes in Kaart rekenen we voor de extra uitgaven aan het Zorginstituut met €30 miljoen per jaar. Dit betekent een forse groei ten opzichte van de huidige rijksbijdrage aan het Zorginstituut van totaal €60 miljoen: de huidige structurele rijksbijdrage is circa €45 miljoen en de incidentele rijksbijdrage circa €15 miljoen (Zorginstituut, 2020). Ondanks dat de voorgestelde maatregelen voor effectief pakketbeheer veel vragen van het Zorginstituut, verwachten we daarom dat dit bedrag ruim voldoende is om deze taken uit te voeren. Het is ook niet realistisch een veel grotere groei van het Zorginstituut binnen enkele jaren te verwachten.

De investeringen in de onderzoeksinfrastructuur en de kosten per evaluatieonderzoek hangen sterk met elkaar samen. We werken met drie scenario's, die we onderaan deze paragraaf verder toelichten:

- Bepaalde infrastructuur: €360.000 per onderzoek per jaar.
- Midden infrastructuur: €180.000 per onderzoek per jaar.
- Sterke infrastructuur: €90.000 per onderzoek per jaar.

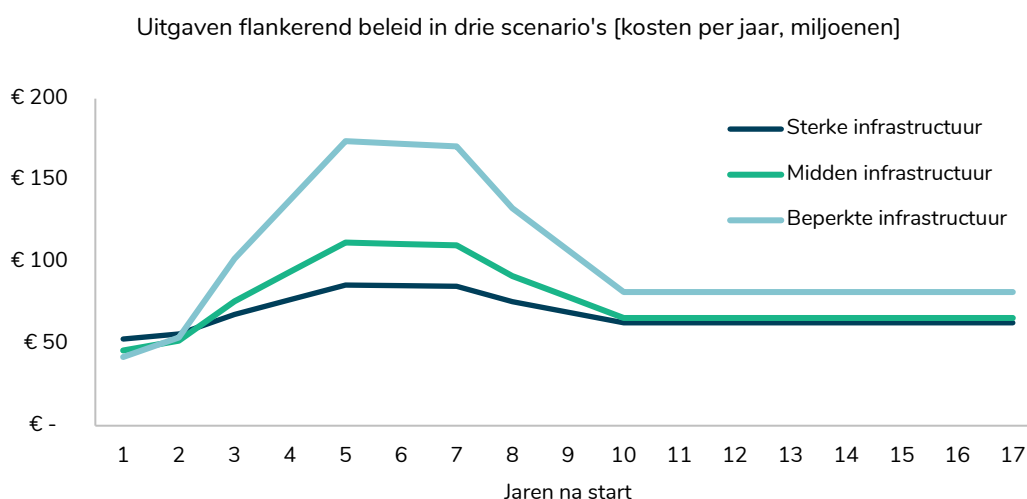
In deze paragraaf nemen we de kosten voor de geleverde zorg (in onderzoeksverband) niet mee in de berekende onderzoekskosten. Dat sluit aan bij het feit dat we de besparingen in de msz-uitgaven ook pas meenemen vanaf de implementatiefase, nadat het onderzoek is afgerond.

Om de jaarlijkse kosten uit te rekenen, gaan we uit van circa 500 benodigde onderzoeken om de niet-effectieve msz te identificeren. Dat aantal is gebaseerd op de aanname dat er enkele honderden niet-effectieve behandelingen geïdentificeerd moeten worden om het besparingspotentieel te realiseren. Deze orde van grootte volgt bijvoorbeeld uit een inventarisatie van de top 10 handelingsagenda van het programma ZE&GG, waaruit 268 kennisvragen naar voren kwamen. Dit aantal valt ook te schatten uit eerdere berekeningen: uit verschillende studies concluderen we dat de meerkosten van niet-effectieve msz ongeveer 6% bedragen. Uit onze vertaling van (Schwartz & al., 2014) naar Nederland, concluderen we dat circa een tiende daarvan te behalen is op basis van 26 ingrepen. Om de totale 6% te realiseren zouden met eenzelfde

impact dus 260 ingrepen geïdentificeerd moeten worden. Voor het benodigde aantal evaluatieonderzoeken verdubbelen we dit nog eens, ervan uitgaande dat de helft van de evaluatieonderzoeken niet-effectieve zorg aan kan tonen. Dit is iets optimistischer dan de gehanteerde waarden in §2.2.1, omdat we ervan uitgaan dat goede agendering (§4.1.2) de inzet effectiever maakt.

Dit betekent tevens dat voor de instroom van nieuwe zorg en indicatieverruiming gerekend moet worden met ongeveer 25 nieuwe onderzoeken per jaar. Met een doorlooptijd van vijf jaar per onderzoek, betekent dit dus dat er structureel circa 125 onderzoeken per jaar lopen.

Gezamenlijk vertalen deze overwegingen zich in jaarlijkse uitgaven die zes jaar na de start maximaal zijn met circa €100 miljoen per jaar en structureel circa €60 miljoen per jaar bedragen (Figuur 17).



Figuur 17 Uitgaven bedragen zes jaar na de start circa €100 miljoen per jaar en structureel circa €60 miljoen per jaar.

Tot slot merken we op dat deze doorrekeningen gericht zijn op onderzoek in Nederland. We verwachten dat efficiënt gebruik van (uitkomsten van) onderzoeken in het buitenland het aantal benodigde onderzoeken in Nederland omlaag kan brengen. Daarmee kunnen mogelijk met lagere jaarlijkse uitgaven vergelijkbare resultaten geboekt worden.

Beperkte infrastructuur: €360.000 per onderzoek per jaar

Voor het scenario met beperkte infrastructuur gaan we uit van Zorgkeuzes in Kaart, dat geen extra infrastructurale investeringen voorziet en uitgaat van €360.000 per onderzoek per jaar (= €90 miljoen voor 50 onderzoeken met doorlooptijd van vijf jaar).

Midden infrastructuur: €180.000 per onderzoek per jaar

Het midden scenario is gekozen tussen beperkt en sterk.

Sterke infrastructuur: €90.000 per onderzoek per jaar

Voor scenario met sterke infrastructuur gaan we uit van €90.000 per onderzoek per jaar. Dit bedrag komt grofweg overeen met de kosten van projecten uit het ZonMw DO programma. Uit de

evaluatie daarvan blijkt dat tussen 2006 en 2017 €139 miljoen is uitgegeven voor 308 projecten (SiRM, 2018). Andermaal uitgaande van een doorlooptijd van vijf jaar per project, resulteert dit in €90.000 per onderzoek per jaar (= €139 miljoen / 308 / 5).

We nemen aan dat om dit als gemiddelde voor evaluatieonderzoeken te bereiken een extra investering in de onderzoeksinfrastructuur nodig is van ruwweg €20 miljoen per jaar. Als referentiewaarde voor dit bedrag kan bijvoorbeeld gekeken worden naar de jaarlijkse uitgaven van IKNL – de grootste registratie van kwaliteitsinformatie en een belangrijke bron voor onderzoek naar effectiviteit van oncologische behandelingen – à €40 miljoen per jaar. Voor een deel bestaat dit uit dataverzameling en onderzoek, wat in ons scenario onder de variabele kosten valt. Op basis van de personele formatie volgens (IKNL, 2017) schatten we dat circa twee derde van de kosten hieraan toe te schrijven is. We rekenen daarom met de overige kosten (een derde) à €10 tot €15 miljoen. Voor een landelijke onderzoeksinfrastructuur is vermoedelijk een hoger bedrag nodig, maar het is onduidelijk in hoeverre huidige gefragmenteerde uitgaven hiervoor gebundeld kunnen worden en in hoeverre daadwerkelijk extra investeringen nodig zijn.

Andere voorbeelden van organisaties waar gegevens systematisch worden verzameld zijn DHD en het Dutch Institute for Clinical Auditing. Voor deze organisaties zijn geen publieke gegevens over de kosten bekend. IKNL, DHD als Dica beschikken waarschijnlijk al over een deel van de benodigde infrastructuur.

Bijlage I Bibliografie

- Badgery-Parker, & al. (2018). Low-value care in Australian public hospitals: prevalence and trends over time.
- Berwick, D., & Hackbarth, A. (2012). Eliminating Waste in US Health Care. *Journal of the American Medical Association*, 1513.
- Bhatia, R., W, L., & Shortt S, e. a. (2015). Measuring the effect of Choosing Wisely: an integrated framework to assess campaign impact on low-value care. *BMJ Qual Saf*, 523-531.
- BMJ. (2012). What conclusions has Clinical Evidence drawn about what works, what doesn't based on randomised controlled trial evidence? Retrieved from <https://web.archive.org/web/20160414100612/http://clinicalevidence.bmj.com/x/set/static/cms/efficacy-categorisations.html>
- Carrol, A. (2017). The High Costs of Unnecessary Care (Jama Forum). *Journal of the American Medical Association* , 1748-1749.
- Centraal Planbureau. (2019). *Middellange termijnverkenning zorg 2022-2025*.
- Chernew, M., & Newhouse, J. (2012). Health Care Spending Growth. In *Handbook of Health Economics*. Elseiver.
- CPB. (2017). *De introductie van dure technologie in de zorg*.
- CPB. (2020). *Evaluatie programma's Beatrixziekenhuis en Bernhoven*.
- Dunn, A., Whitmer, B., Batch, A., Fernando, L., & Rittmueler, L. (2018). High Spending Growth Rates for Key Diseases In 2000-14. *Health Affairs*.
- Eén vandaag opiniepanel. (2015). *Onderzoek: effectiviteit van zorg*.
- Europese Commissie. (2010). *The role of technology in healthcare expenditure in the EU*. Brussel: DG Economic and Financial Affairs.
- FMS. (2016). *Adviesrapport Zorgevaluatie van project naar proces*.
- Garner, S., & Littlejohns, P. (2011). Disinvestment from low value clinical interventions:. *British Medical Journal*, 343.
- Garrow, J. (2007). *How much of orthodox medicine is evidence based*.
- Heida, J., Den Engelsen, B., Baeten, S., & Van Gent, C. (2016). *Productmarktafbakening ziekenhuiszorg*. Den Haag.
- Herrera-Perez, & al. (2019). A comprehensive review of randomized clinical trials in three medical journals reveals 396 medical reversals.
- IKNL. (2017). *Meerjarenbeleidsplan 2018-2022*.

- Imrie, & Ramey. (2000). *The evidence for evidence based medicine*.
- Institute of Medicine. (2010). *The healthcare imperative: lowering costs and improving outcomes: workshop series summary*. Washington: The National Academies.
- Institute of Medicine. (2013). *Best Care at Lower Cost: The path to continuously learning health care in America*.
- KPMG Plexus. (2010). *Reduceren van praktijkvariatie: budgettaire effecten van scherpere indicatiestelling*.
- Norton, W., & Chambers, D. A. (2020). Unpacking the complexities of de-impelmenting inappropriate health interventions. *Implementation science*, 15:2.
- Norton, W., Kennedy, A., & Chambers, D. (2017). Studying de-implementation in health: an analysis of funded research grants. *Implementation Research* , 12:144.
- OECD. (2017). *Tackling wasteful spending on health*.
- Prasad, V., Vandross, A., Toomey, C., Cheung, M., Rho, J., Quinn, S., . . . Cify, A. (2013, August). A Decade of Reversal: An Analysis of 146 contradicted medical practices. *Mayo Clinic Proceedings*, 790-798.
- PwC Health Research Institute. (2010). *The price of excess: Identifying waste in healthcare spending*.
- Repping, S. (2019). *Rapportage Kwartiermakersfase Zorgevaluatie en Gepast Gebruik*.
- Schwartz, & al. (2014). *Measuring low-value care in medicare*.
- SiRM. (2018). *Benut potentieel - Evaluatie programma DoelmatigheidsOnderzoek 2006-2017*.
- SiRM. (2019). *Aanbevelingen ter bevordering van gelijkgerichtheid*.
- Sorensen, C., Drummond, M., & Bhuiyan Khan, B. (2013). Medical technology as a key driver of rising health. *ClinicoEconomics and Outcome Research*, 223-234.
- Tovey, D. (2007, november 10). Evidence should not be the only consideration: patient's and clinicians' views matter too. *BMJ*, p. 952.
- Wammes, & al. (2016). *Identifying and prioritizing lower value services from Dutch specialist guidelines and a comparison with the UK do-not-do list*.
- Zorginstituut. (2020). *Jaarverslag 2019 Zorginstituut Nederland*.

Bijlage 2 Interviews en werksessies

Informatie voor dit rapport is mede verzameld door interviews en twee werksessies met deskundigen en betrokkenen bij het onderwerp (Tabel 18).

Naam	Organisatie	Interview	Werksessie 1	Werksessie 2
Bas Jurling	NZa	V	V	
Erik Langeveld	NZa	V		
Gerjanne Vianen	ZonMw	V		V
Joris Müskens	Doen of laten / IQ Healthcare	V	V	
Lonneke Timmers	Zorginstituut	V		V
Mariëlle Snijders	ZonMw	V		V
Matthijs Versteegh	iMTA	V		V
Patrick Jeurissen	Radboudumc	V	V	
Pauline de Heer	Zorginstituut	V	V	
Sjoerd Repping	ZE&GG	V		V
Teus van Barneveld	FMS	V		V
Tijmen Oegema	IGJ	V		
Tijn Kool	Doen of laten / IQ Healthcare	V	V	
Vera Luggenhorst	NZa		V	V

Tabel 18 Er zijn 13 personen geïnterviewd en 12 personen hebben bijgedragen aan een van de twee werksessies.