



Cochrane
Netherlands

Van black-box naar white-box: beoordeling en evaluatie van digitale zorgtechnologieën

Eindrapportage

27 mei 2024

Kevin Jenniskens
Pauline Heus
Linde Huis in 't Veld
Demy Idema
Mike Kusters
René Spijker
Maarten van der Laan
Lotty Hooft



Van black-box naar white-box beoordeling en evaluatie van innovaties

Academische werkplaats in samenwerking met Zorginstituut Nederland (ZIN) en het consortium Kwaliteit van Zorg van de Nederlandse Federatie van Universitair Medische Centra (NFU)

Samenstelling projectteam:

Onderzoekers

- Dr. K. Jenniskens
- Dr. P. Heus
- Drs. L. Huis in 't Veld
- Drs. D. L. Idema
- Drs. M. Kusters
- Prof. Dr. L. Hooft

Informatiespecialist

- Drs. R. Spijker

Betrokkenen ZIN

- Dr. P. Abrishami
- Dr. Ir. A. Chorus
- Drs. R.M. Toorneman
- Dr. S. Kleijnen,

Betrokkene NFU

- Dr. M.J. van der Laan

Contactgegevens

Dr. K. Jenniskens
Cochrane Netherlands / Julius Centrum
Universitair Medisch Centrum Utrecht
Huispostnummer Str. 6.131
Postbus 85500
3508 GA Utrecht
T: 088 755 9301
E: cochrane@umcutrecht.nl
W: www.cochrane.nl

Disclaimer: The views and opinions expressed in this document are those of the authors and do not necessarily reflect those of The Cochrane Collaboration.

Lijst met gebruikte afkortingen

BfArM	Bundesinstitut für Arzneimittel und Medizinprodukte (Duitsland)
DHT	Digital health technologies (digitale zorgtechnologieën)
ESF	Evidence Standards Framework
EUnetHTA	European Network of Health Technology Assessment
FDA	Food & Drug Administration (Verenigde Staten)
GoeG	Gesundheit Österreich GmbH (nationale gezondheidszorginstituut Oostenrijk)
HI-NL	Health Innovation Netherlands
HTA	Health technology assessment (economisch evaluatieonderzoek in de gezondheidszorg)
INHTA	International Network of Agencies for Health Technology Assessment
IZA	Integraal Zorgakkoord (IZA)
KCE	Federaal Kenniscentrum voor de gezondheidszorg (België)
MDR	Medical Device Regulation (Europese verordening voor medische hulpmiddelen)
MIB	MedTech innovation briefing
NFU	Nederlandse Federatie van Universitair Medische Centra
NHS	National Health Service (Verenigd Koninkrijk)
NICE	National Institute of Clinical Excellence (Verenigd Koninkrijk)
RCT	Randomized controlled trial (gerandomiseerd gecontroleerd onderzoek)
RWE	Real-world evidence
SWP	Stand van Wetenschap en Praktijk
UMC	Universitair Medisch Centrum
WLZ	Wet langdurige zorg
ZIN	Zorginstituut Nederland
ZVW	Zorgverzekeringswet

Begrippenlijst

Om de resultaten uit deze rapportage op een juiste manier te kunnen interpreteren en correct in context te kunnen plaatsen, is het van belang om specifieke begrippen expliciet te definiëren. Daartoe is hieronder een begrippenlijst met bijbehorende definities te vinden, voor enkele termen die binnen dit rapport worden gehanteerd.

Door het gebrek aan consensus omtrent deze definities in de wetenschappelijke literatuur, is het geven van sluitende definities vrijwel onmogelijk. Deze begrippenlijst poogt enkel, met behulp van enkele voorbeelden en specifieke duidingen, een kapstok te bieden waaraan de resultaten van deze rapportage kunnen worden opgehangen.

Digitale zorg(innovatie)	De toepassing van zowel digitale informatie als communicatie om de gezondheid en gezondheidszorg te ondersteunen en/of te verbeteren
Plegen te bieden criterium	Toetst of een interventie* behoort tot zorg die de beroepsgroep rekent tot het aanvaarde arsenaal van zorg, en die geleverd wordt op een wijze die de betreffende beroepsgroep als professioneel juist beschouwt.
Beoordeling Stand van de Wetenschap en Praktijk (SWP)	Evaluatieproces waarin wordt beoordeeld of een digitale interventie bij een bepaalde indicatie voldoende meerwaarde of gelijke waarde heeft ten opzichte van de standaard- of gebruikelijke behandeling. Enkel interventies waarvan ten minste is vastgesteld dat ze geen technische variant zijn, en waarvan wordt vastgesteld dat ze voldoen aan het 'plegen te bieden' criterium, worden onderworpen aan de beoordeling in het kader van SWP. Naast de conclusie omtrent SWP wordt in sommige gevallen een maatschappelijke weging (<i>appraisal</i>) van andere pakketcriteria (d.w.z. kosteneffectiviteit, noodzakelijkheid en uitvoerbaarheid) toegepast. Dit resulteert in een pakketadvies aan de minister van VWS.
Technische variant (beoordeling Stand van de Wetenschap en Praktijk)	Toetst of een (digitale) interventie een variant is van een interventie die reeds beschikbaar is in de klinische praktijk, en of deze in het basispakket opgenomen is. Indien wordt verondersteld dat een digitale interventie een technische variant is op een bestaande interventie, kan mogelijk worden aangenomen dat de effectiviteit hiervan niet wezenlijk afwijkt.
Verklarende RCT	Traditionele RCT, waarbij het primaire doel is om te onderzoeken of een digitale zorginnovatie kan werken onder ideale omstandigheden. Deze onderzoeksopzet heeft onder andere strikte in- en exclusiecriteria, vaardigheden van zorgpersoneel, en hoge therapietrouw van studiedeelnemers.
Pragmatische RCT	RCT met een aangepaste onderzoeksopzet, waarbij het primaire doel is om te onderzoeken of een digitale interventie kan werken in de dagelijkse praktijk. Dit onderzoeksopzet biedt meer flexibiliteit t.o.v. de genoemde criteria bij een verklarende RCT.

Retrospectieve studie	Studie waarin data voorafgaand aan het formuleren van een specifieke onderzoeksvraag verzameld zijn. Deze onderzoeksopzet vergt minder tijdsinvestering van onderzoekers, maar heeft grotere kans op ontbrekende of heterogeen gerapporteerde data.
Prospectieve studie	Studie waarin data naar aanleiding van een specifiek geformuleerde onderzoeksvraag verzameld worden. Dit vergt een grotere tijdsinvestering van onderzoekers, maar heeft een grotere kans complete en homogeen gerapporteerde data.
Observationele studie	Studie waarbij data verworven uit de routine klinische praktijk worden gebruikt om een onderzoeksvraag te beantwoorden. Patiënten kunnen enkel worden onderworpen aan digitale zorginnovaties die reeds beschikbaar zijn in de klinische praktijk.
Real-world Evidence (RWE)	Data die verzameld zijn of zullen worden uit de dagelijkse klinische praktijk. RWE worden voor het overgrote merendeel verkregen uit observationeel onderzoek en kent daarmee vergelijkbare uitdagingen. Hoewel pragmatische trials beter in staat zijn de klinische praktijk te reflecteren, worden deze niet gezien als RWE, omdat dit inherent nog steeds een experimentele onderzoeksopzet is.
GRADE-methodiek	Methode om op een transparante wijze tot een oordeel te komen over de kwaliteit van het bewijs (bewijskracht t.a.v. de effectiviteit) en de sterkte van de daaruit volgende aanbeveling.

* Het begrip interventie kan breed worden opgevat; behandeling, diagnostiek, predictiemodellen, (digitale) zorginnovaties, etc

Inhoudsopgave

Samenvatting	7
1. Introductie	11
2. Methoden	14
2.1 <i>Verkennend literatuuronderzoek</i>	14
2.1.1 <i>Beoordelingscriteria</i>	15
2.1.2 <i>Beoordelingsprocedure</i>	16
2.2 <i>Kwalitatief onderzoek</i>	17
2.2.1 <i>Focusgroepen</i>	17
2.2.2 <i>Vragenlijst</i>	18
2.3 <i>Logic model</i>	19
3. Resultaten	20
3.1 <i>Verkennend literatuuronderzoek</i>	20
3.1.1 <i>Beoordelingscriteria</i>	20
3.1.2 <i>Beoordelingsprocedure</i>	25
3.2 <i>Kwalitatief onderzoek</i>	32
3.2.1 <i>Focusgroepen</i>	32
3.2.2 <i>Vragenlijst</i>	36
3.2.2.1 <i>Resultaten gesloten vragen</i>	36
3.2.2.2 <i>Resultaten open vragen</i>	37
3.3 <i>Logic model</i>	47
4. Discussie	50
Bijlagen	52
<i>Bijlage 1. Details focusgroepen</i>	52
<i>Bijlage 2. Opzet vragenlijst</i>	54
<i>Bijlage 3. Flowchart verkennend literatuuronderzoek</i>	60
<i>Bijlage 4. Beschrijving beoordelingsprocedure per instantie en land</i>	61
<i>Bijlage 5. Kwantitatieve analyse vragenlijst: subgroep o.b.v. affiliatie</i>	85
<i>Bijlage 6. Kwalitatieve analyse vragenlijst</i>	87

Samenvatting

Binnen de gezondheidszorg is er in toenemende mate aandacht voor digitale zorginnovaties. Een digitale zorginnovatie kan worden omschreven als de toepassing van zowel digitale informatie als communicatie om de gezondheid en gezondheidszorg te ondersteunen en/of te verbeteren. Voorbeelden van digitale zorginnovaties zijn apps of toepassingen die patiënten en zorgverleners gebruiken om informatie digitaal te verzamelen, registreren, analyseren, uit te wisselen of te communiceren.

Digitale zorginnovaties zijn essentieel om de doelstellingen te behalen die beschreven worden in het Integraal Zorgakkoord (IZA). De toenemende inzet van deze innovaties brengt vraagstukken mee op het gebied van de evaluatie ervan. Alle vergoede zorg (zowel nieuwe als bestaande zorg) binnen de Zorgverzekeringswet (Zvw) of Wet langdurige zorg (Wlz) moet voldoen aan de wettelijke criteria 'Plegen te bieden' en 'Stand van Wetenschap en Praktijk' (SWP). Op het niveau van een individuele verzekerde geldt bovendien het 'Redelijkerwijs aangewezen op' criterium. Derhalve zijn deze criteria dus ook van toepassing op digitale zorg.

Randomized controlled trials (RCTs) worden gezien als de gouden standaard om effectiviteit van een zorgvernieuwing of innovatie aan te tonen. Echter is een dergelijke onderzoeksopzet niet altijd haalbaar, wenselijk of noodzakelijk. Er bestaat in Nederland geen formeel raamwerk of consensus over wanneer welke alternatieve onderzoeksopzetten en methoden een compleet evidenceportfolio kunnen genereren met voldoende bewijskracht voor het aantonen van effectiviteit van een innovatie.

De beschreven uitdagingen zijn de aanleiding en basis voor deze academische werkplaats "Van black-box naar white-box: beoordeling en evaluatie van innovaties", die in samenwerking met de Nederlandse Federatie van Universitair Medische Centra (NFU), Zorginstituut Nederland (ZIN), Cochrane Netherlands en het Universitair Medisch Centrum (UMC) Utrecht tot stand is gekomen. Het doel van deze academische werkplaats is om een raamwerk van beoordelingscriteria en bijpassende onderzoeksopzetten te ontwikkelen als basis voor toekomstige uitwerking en implementatie in een Nederlands beoordelingskader.

Om dit doel te bereiken werden de volgende drie stappen uitgevoerd:

- Een verkennend literatuur- en richtlijnenonderzoek (*scoping review*), dat een overzicht geeft van nationale en internationale beoordelingscriteria, methoden, en analyseraamwerken die gebruikt worden voor evaluatie en toetsing/beoordeling van digitale zorginnovaties.
- Kwalitatief onderzoek, waarin de resultaten van het literatuuronderzoek gebruikt worden als basis voor verdiepende focusgroepen, en daaropvolgend een vragenlijstonderzoek onder een bredere groep stakeholders wordt uitgevoerd.
- Het creëren van een *logic model*, gebaseerd op de voornaamste bevindingen uit het verkennende literatuuronderzoek en het kwalitatieve onderzoek, dat een schematisch overzicht geeft van uitdagingen, behoeften, gewenste resultaten, mogelijke strategieën en onderliggende voorwaarden/assumpties.

Uit het literatuur- en kwalitatieve onderzoek kwamen verscheidene bevindingen naar voren. Zo bleek in het kwalitatieve onderzoek onduidelijkheid te bestaan over welke digitale zorginnovaties een (volledige) beoordeling SWP moeten (of kunnen) doorlopen. Enerzijds omdat er onduidelijkheid is rondom het ‘plegen te bieden’ criterium, maar ook omdat het inschatten van de effectiviteitsvraag uitdagend kan zijn: wanneer kan een digitale zorginnovatie bijvoorbeeld gezien worden als een technische variant, en wanneer is er sprake van een nieuwe interventie?

Daarnaast was één van de bevindingen in het verkennende literatuuronderzoek dat er onderscheid werd gemaakt tussen typen digitale zorginnovaties op basis van risicoprofielen en het doel van inzet van de innovatie. In het kwalitatieve onderzoek werd eveneens geconstateerd dat het risico op nadelige gevolgen voor de gezondheid niet hetzelfde is voor alle typen digitale zorginnovaties, denk aan inzet van *telemonitoring* ten opzichte van de inzet van *digital therapeutics*.

Het benodigde evidenceportfolio moet passend zijn voor het type digitale zorg dat geëvalueerd dient te worden. Bij een laag-risicoprofiel zou bewijs van lagere kwaliteit of minder bewijs acceptabel kunnen zijn, en een RCT niet noodzakelijk. Er is een scala aan alternatieve onderzoeksopzetten genoemd, variërend van RCTs in aangepaste vorm tot observationele studies en gebruikmaken van *real world evidence (RWE)*. Digitale zorginnovaties zijn veelal geschikt om grote hoeveelheden (RWE) data te verzamelen. Monitoring wordt gezien als een mogelijk alternatief voor situaties waarin een RCT niet haalbaar is: de innovatie al toepassen in de praktijk (en vergoeden) en parallel daaraan data verzamelen om effectiviteit (en andere relevante uitkomsten) te kunnen onderzoeken. Onvoldoende (kwaliteit van) dataregistratie en uitdagingen om causale relaties aan te tonen worden als belangrijke barrières gezien.

Uit het kwalitatieve onderzoek kwam ook naar voren dat veel van de claims bij digitale zorginnovaties zich niet (alleen) richten op effectiviteit, maar bijvoorbeeld (ook) op doelmatigheid, organisatie of continuïteit van zorg. Het gaat dan over pakketbeheer in de breedte. Het feit dat de overige pakketcriteria (kosteneffectiviteit, noodzakelijkheid en uitvoerbaarheid) alleen in bepaalde gevallen beoordeeld worden en enkel als gezondheidswinst is aangetoond, werd als belemmering genoemd. Duurzaamheid en arbeidsinzet worden gezien als waardevolle aanvullende criteria die voor de brede maatschappelijke opgave relevant zijn.

De bevindingen uit deze academische werkplaats geven richting voor mogelijke vervolgstappen. Er is behoefte aan duidelijkheid rondom onderwerpen in beoordelingsproces zoals welke digitale zorginnovaties een (volledige) beoordeling SWP dienen te doorlopen, voorbeelden rondom het ‘plegen te bieden’ criterium en ‘technische variant’. Daarnaast zouden niet alle digitale zorginnovaties eenzelfde impact op gezondheid hebben, waarmee met betrekking tot het benodigde evidenceportfolio gedacht zou kunnen worden aan stratificatie op basis van risicoprofielen, gerelateerd aan type innovatie. Er wordt echter ook gepleit voor het toekennen van meer aandacht en gewicht aan de overige overwegingen naast de beoordeling van effectiviteit. Dit past bij de maatschappelijke uitdagingen omtrent druk op de zorg en zorgprofessionals. Bij implementatie dient rekening te worden gehouden met een breder pallet aan factoren zoals schaalbaarheid van de

innovatie, duurzame inzet, arbeidsinzet en aan welke informatievoorziening en ondersteuning innovatoren behoefte hebben.

De bevindingen uit deze academische werkplaats kunnen als basis dienen voor toekomstige projecten in het kader van de evaluatie van digitale zorginnovaties. Verdere uitwerking van het kader kan zorgprofessionals, wetenschappers, beleidsmakers en het brede publiek transparantie en helderheid verschaffen en innovatoren helpen bij het verzamelen van passend bewijs voor een evidenceportfolio van hun innovatie.

De belangrijkste (beleids)aanbevelingen in één oogopslag

Naar aanleiding van de bevindingen uit deze academische werkplaats zijn de volgende aanbevelingen geformuleerd om richting te geven aan toekomstig beleid van Zorginstituut Nederland:

- Verduidelijk hoe het criterium ‘plegen te bieden’ dient te worden toegepast op digitale zorg en hoe moet worden gehandeld in geval een digitale zorginnovatie (nog) geen onderdeel is van de professionele (behandel)richtlijnen van de beroepsgroep.
- Verduidelijk met behulp van voorbeelden wanneer digitale zorg een ‘technische variant’ is. Vergroot de bekendheid van het begrip technische variant binnen het beoordelingskader SWP, opdat veldpartijen zelfstandig deze toetsing kunnen toepassen.
- Verduidelijk welke verschillende typen digitale zorg kunnen worden onderscheiden en wanneer een beoordeling SWP aan de orde is. Overweeg hierbij risicostratificatie naar typen (bijv. a.d.h.v. CE-risicoklassen) en pas de gevraagde zwaarte/kwaliteit van bewijs hierop aan.
- Ontwikkel een visie op het gebruik van alternatieve onderzoeksopzetten en RWE in het kader van deze typen digitale zorginnovaties. Ontwikkel de juiste beleidsinstrumenten waarmee veldpartijen en ontwikkelaars kunnen inschatten welke onderzoeksopzetten kunnen worden overwogen bij verschillende typen digitale zorg(innovaties) en biedt veldpartijen handvatten om dit instrument zelf toe te passen.
- Overweeg een versie van de passend onderzoek-/bewijsvragenlijst te ontwikkelen waarin het taalgebruik meer is afgestemd op digitale zorg.
- Verduidelijk in hoeverre digitale zorg tijdens een klinische studie of na instroom in het verzekerde pakket nog mag worden doorontwikkeld of uitgebreid (i.e. anders dan technische/veiligheidsupdates). Verduidelijk ook in hoeverre de context waarin de digitale zorg wordt toegepast consequenties heeft voor het effectiviteitsbewijs of voor vergoeding.
- Overweeg in het licht van de maatschappelijke opgaven of/hoe in de toekomst overige overwegingen (d.w.z. kosteneffectiviteit, noodzakelijkheid, uitvoerbaarheid, maar ook bijv. duurzaamheid en arbeidsinzet) op een meer gestandaardiseerde manier kunnen worden meegewogen binnen het huidige wettelijke kader.
- In verschillende landen kunnen ontwikkelaars van digitale zorg zich direct wenden tot een formele vergoedingsinstantie. In Nederland kan dit in zeer beperkte mate of niet. Onderzoek wat de impact hiervan is op de instroom van digitale zorginnovaties.

- Overweeg nader onderzoek naar de noodzaak en mogelijkheid om een gezamenlijke bekostiging of (subsidie)gelden beschikbaar te stellen om voor startups de drempel voor het verrichten van (effectiviteits)onderzoek naar digitale zorginnovaties te verlagen, vergelijkbaar met de 'Veelbelovende Zorg' (VeZo) en STOZ subsidies.
- Er is maatschappelijke urgentie voor transparantie en standaardisering van de evaluatie van digitale zorginnovaties. Waardering van passend bewijs kan niet (volledig) aan de markt/veldpartijen worden gelaten, derhalve verdient het de overweging t.a.v. het Zorginstituut om hier een regierol in te pakken.

1. Introductie

Binnen de gezondheidszorg is er in toenemende mate aandacht voor digitale zorginnovaties (ook wel digitale zorgtechnologieën of *digital health technologies* (DHT) in internationale literatuur). Er bestaan veel definities van digitale zorginnovaties. In dit rapport hanteren wij de definitie van Nictiz uit 2019¹, waarin een digitale zorginnovatie wordt omschreven als de toepassing van zowel digitale informatie als communicatie om de gezondheid en gezondheidszorg te ondersteunen en/of te verbeteren. Deze brede definitie geeft aan dat een digitale zorginnovatie meer is dan enkel de techniek: het gaat om het toepassen ervan om een hoger doel op het gebied van gezondheid en gezondheidszorg te bereiken. Voorbeelden van digitale zorginnovaties zijn apps of toepassingen die patiënten en zorgverleners gebruiken om informatie digitaal te verzamelen, registreren, analyseren, uit te wisselen of te communiceren. Die digitale informatie kan vervolgens worden gebruikt voor bijvoorbeeld beslisondersteuning, zorg op afstand, onderzoek of doeleinden op het gebied van kwaliteitsverbetering.

Digitale zorginnovaties zijn essentieel om de doelstellingen te behalen die beschreven worden in het Integraal Zorgakkoord (IZA).² Naar verwachting zullen deze innovaties dan ook steeds meer ingezet worden, waarbij, in het kader van passende zorg en de wettelijke criteria voor verzekerde zorg, alleen de juiste digitale zorginnovaties dienen te worden opgeschaald. Dit brengt vraagstukken mee op het gebied van evaluatie van digitale zorginnovaties: welke *evidence* is voldoende om de effectiviteit, kwaliteit, en haalbaarheid te waarborgen? En welke onderzoeksopties passen het beste bij een bepaald type *evidence*? Antwoorden op deze vragen zijn een belangrijke stap in het stimuleren van digitale transitie conform *Passende Zorg* principes en daarmee in het bereiken van deze doelstelling van het IZA.

Voor vergoeding vanuit de Zorgverzekeringswet (Zvw) of Wet langdurige zorg (Wlz) is het allereerst van belang dat een innovatie ingezet wordt ondersteunend is binnen een geneeskundige/medische context. Alle vergoede zorg (zowel nieuwe als bestaande zorg) binnen de Zvw en de Wlz moet voldoen aan de wettelijke criteria ‘Plegen te bieden’ en ‘Stand van de wetenschap en praktijk’ (SWP).^{3,4} Deze criteria zijn dus ook van toepassing op digitale zorg. Zorg die een beroepsgroep ‘pleegt te bieden’ betreft zorg die een beroepsgroep rekent tot het aanvaarde arsenaal van zorg, en die geleverd wordt op een wijze die de beroepsgroep beschouwt als professioneel juist. Daarnaast dient op het niveau van een individuele verzekerde te worden voldaan aan het ‘redelijkerwijs aangewezen op’ criterium.

¹ E-health, wat is dat? Nictiz. Published March 27, 2019. Accessed April 19, 2023. <https://nictiz.nl/publicaties/e-health-wat-is-dat/>

² Ministerie van VWS. Integraal Zorgakkoord: “Samen werken aan gezonde zorg” - Rapport - Rijksoverheid.nl. Published September 16, 2022. Accessed April 8, 2024. <https://www.rijksoverheid.nl/documenten/rapporten/2022/09/16/integraal-zorgakkoord-samen-werken-aan-gezonde-zorg>

³ <https://www.zorginstituutnederland.nl/publicaties/rapport/2008/11/17/betekenis-en-beoordeling-criterium-plegen-te-bieden>

⁴ Ministerie van VWS. Beoordeling stand van de wetenschap en praktijk 2023 - Publicatie - Zorginstituut Nederland. Published April 11, 2023. Accessed April 19, 2023. <https://www.zorginstituutnederland.nl/publicaties/publicatie/2023/04/11/beoordeling-swp-2023>

Aan de hand van het beoordelingskader SWP wordt bepaald of zorg op groepsniveau in voldoende mate bewezen effectief is conform de principes van *evidence-based medicine*. Hierbij wordt nagegaan of een interventie (diagnostiek, behandeling) resulteert in een relevante (meer)waarde voor de doelgroep ten opzichte van de standaardbehandeling. Er moet sprake zijn van nettowinst op uitkomstmaten die voor de doelgroep/patiënt relevant zijn. Binnen het beoordelingskader SWP wordt voorafgaand aan het verzamelen van literatuur vastgesteld in samenspraak met de betrokken partijen wat de plaatsbepaling is van de interventie in het zorgproces van een patiëntengroep en wat de interventie beoogt. Vervolgens wordt een inschatting gemaakt van welke studieopzet wenselijk en haalbaar is.⁵

Randomized controlled trials (RCTs) worden gezien als de gouden standaard om effectiviteit van een zorgvernieuwing of innovatie te onderzoeken. Indien goed uitgevoerd levert dit type onderzoeksoopzet bewijs van de hoogste kwaliteit. Tegelijkertijd is dit type onderzoeksoopzet is lang niet altijd haalbaar, wenselijk of noodzakelijk. Cochrane Netherlands heeft daartoe op verzoek van ZonMW in mei 2013 een kennissynthese uitgevoerd over “*Alternatieven voor RCTs in onderzoek naar de effectiviteit van interventies*”, en hierbij deze alternatieve onderzoeksoopzetten in kaart gebracht.⁶

De ervaring leert echter dat er onder onderzoekers en innovatoren nog steeds onduidelijkheid bestaat over de waarde van alternatieve onderzoeksoopzetten en hoe deze kunnen worden ingezet voor evaluatie van de relatie tussen de context waarin een zorginnovatie wordt toegepast, mechanismen die teweeg worden gebracht door de innovatie, en de impact op behaalde uitkomsten.

Landelijke initiatieven, zoals Health Innovation Netherlands (HI-NL)⁷, trachten innovatieonderzoekers te ondersteunen bij het ontwerpen van de onderzoeksoopzet passend bij de verschillende fasen uit de ontwikkeling en evaluatie van de impact van een zorginnovatie. Echter, ook binnen HI-NL is er – na evaluatie van meer dan 20 zorginnovaties – tot op heden nog geen hiërarchisch overzicht van alternatieve onderzoeksmethoden welke gebruikt kunnen worden bij de evaluatie van een zorginnovatie tijdens de verschillende fasen van de ontwikkeling.

Er blijkt zowel bij Nederlandse beoordelende instanties als bij andere landelijke initiatieven geen formeel raamwerk te zijn of consensus te bestaan over wanneer welke alternatieve onderzoeksmethoden een compleet *evidenceportfolio* met voldoende bewijslast voor een innovatie genereren. Om patiënten eerder toegang te kunnen geven tot digitale zorginnovaties is het van belang helderheid te verschaffen over de meest passende onderzoeksoopzet en de benodigde hoeveelheid bewijslast voor het aantonen van de (klinische en economische) meerwaarde van digitale zorginnovaties t.b.v. acceptatie, implementatie en bekostiging van deze innovaties. Uiteraard blijft het noodzakelijk dat de *evidence* voldoende basis biedt om de effectiviteit en veiligheid van de digitale

⁵ Heymans JM. Passend bewijs bij het bepalen van effectiviteit van interventies. Ned Tijdschr Geneeskd.

⁶ May A, Matthijsen J. Alternatieven voor RCT bij de evaluatie van effectiviteit van interventies!? Mulier Instituut. Published 2015. Accessed March 24, 2023.

<https://www.mulierinstituut.nl/publicaties/24999/alternatieven-voor-rct-bij-de-evaluatie-van-effectiviteit-van-interventies/>

⁷ Bringing Health Innovation to the next level | Health Innovation Netherlands. Published March 4, 2024. Accessed May 1, 2024. <https://www.healthinnovation.nl/bringing-health-innovation-next-level>

zorginnovaties te waarborgen. Er bestaan in Nederland echter nog geen eenduidige criteria waarmee *evidenceportfolio's* verkregen via alternatieve onderzoeksmethoden moeten worden beoordeeld om in aanmerking te 'omen voor implementatie of bekostiging.

De beschreven uitdagingen vormden de aanleiding en basis voor de academische werkplaats “Van black-box naar white-box beoordeling en evaluatie van innovaties”, die in samenwerking met de Nederlandse Federatie van Universitair Medische Centra (NFU), Zorginstituut Nederland (ZIN), Cochrane Netherlands en het Universitair Medisch Centrum (UMC) Utrecht tot stand is gekomen. Het doel van deze academische werkplaats is om een raamwerk van beoordelingscriteria en bijpassende onderzoeksozpetten te ontwikkelen als basis voor toekomstige uitwerking en implementatie in een Nederlands beoordelingskader. Om dit doel te bereiken zijn drie stappen genomen:

1. Het creëren van een systematisch overzicht van beoordelingscriteria, onderzoeksozpetten, en beoordelingsprocessen die momenteel door verschillende (inter)nationale beoordelende instanties worden gehanteerd, en in wetenschappelijke literatuur worden beschreven, ten behoeve van de implementatie en bekostiging van digitale zorginnovaties.
2. Het geven van inzicht in de visie van stakeholders in het Nederlandse zorglandschap ten opzichte van de (inter)nationaal gehanteerde beoordelingscriteria voor implementatie en bekostiging van digitale zorginnovaties, waar zij de meeste waarde aan hechten, en welke belemmerende en bevorderende factoren zij zien ten behoeve van adoptie van digitale zorginnovaties. Een breed gedragen visie over dit onderwerp, inclusief regelgevende belemmeringen en patiëntvertrouwen, is noodzakelijk om het toelaten van een bredere bewijsbasis te ondersteunen en te bevorderen.
3. Het geven van een aanzet tot een analyseraamwerk dat past binnen het huidige beoordelingskader, draagvlak heeft onder alle betrokken stakeholders in Nederland, en door hen gebruikt kan worden als wegwijzer voor de evaluatie en toetsing/beoordeling van zorginnovaties, nu en in de toekomst.

2. Methoden

Deze academische werkplaats is uitgevoerd middels drie opeenvolgende stappen:

- Een **verkennend literatuur- en richtlijnenonderzoek** (*scoping review*), dat een overzicht geeft van nationale en internationale beoordelingscriteria, methoden, en analyseraamwerken die gebruikt worden voor evaluatie en toetsing/beoordeling van digitale zorginnovaties.
- Een **kwalitatief onderzoek**, waarin de resultaten van het literatuuronderzoek werden gebruikt als basis voor verdiepende focusgroepen en daaropvolgend een vragenlijstonderzoek onder een bredere groep stakeholders.
- Het creëren van een **logic model**, gebaseerd op het verkennende literatuuronderzoek en het kwalitatieve onderzoek, dat een schematisch overzicht geeft van uitdagingen, behoeften, gewenste resultaten, strategieën om dit te bereiken en onderliggende voorwaarden/assumpties.

De methodologie behorende bij elk van deze stappen staat in onderstaande secties beschreven.

2.1 Verkennend literatuuronderzoek

Het doel van het verkennende literatuuronderzoek was niet om een alomvattend overzicht van beoordelingscriteria te geven die in alle raamwerken en richtlijnen worden toegepast, maar om een representatieve dataset te genereren die een breed beeld geeft van welke beoordelingscriteria worden beschreven en hoe deze worden ingevuld door de internationale gemeenschap. Dit literatuuroverzicht diende dan ook vooral als basis en startpunt voor verdere discussie in het kwalitatieve gedeelte van dit project.

Bij het verkennende literatuuronderzoek werd een brede zoekstrategie gehanteerd, niet alleen naar wetenschappelijke publicaties, maar ook naar relevante beleidsdocumenten en richtlijnen verstrekt door *health technology assessment (HTA)* organisaties.

Op basis van de literatuur waarnaar gerefereerd werd in een intern visiedocument van ZIN van 16 mei 2022,⁸ werd een database samengesteld met 29 referenties. Vervolgens werd op basis hiervan op 7 november 2022 in Google Scholar gezocht naar publicaties waarin deze referenties aangehaald werden (*cited reference search*). Elk van deze publicaties werd geïncludeerd ten behoeve van verdere selectie.

Selectie van artikelen op basis van titel en abstract werd gedaan door één onderzoeker. Artikelen die een beoordelingskader beschreven dat gebruikt werd voor de evaluatie van digitale zorginnovaties, werden geïncludeerd. Artikelen waarvan de volledige tekst niet vrij toegankelijk was, die niet in Engels of Nederlands waren geschreven, of artikelen die enkel een beoordelingskader voor digitale zorginnovaties ten behoeve van een specifiek ziektebeeld beschreven, werden geëxcludeerd.

Naast wetenschappelijke literatuur werden ook richtlijnen en adviesdocumenten over digitale zorginnovaties van internationale HTA-organisaties bekeken. Hiertoe werd op 7 december 2022

⁸ Zorginstituut Nederland. Intern visiedocument. Published online May 16, 2022.

gezocht naar organisaties die bij het *International Network of Agencies for Health Technology Assessment (INHTA)* en/of het *European Network of Health Technology Assessment (EUnetHTA)* waren aangesloten. Organisaties uit landen in Europa, Noord-Amerika, Japan of Australië werden geselecteerd. Via een gestandaardiseerde zoekopdracht met termen gerelateerd aan eHealth, mHealth, digitale zorg en digitale technologie, werd op de websites van de desbetreffende instanties en ook via Google gezocht naar relevante documenten.

2.1.1 Beoordelingscriteria

Voor de data-extractie van de beoordelingscriteria werd gebruik gemaakt van het *Evidence Standards Framework (ESF)*, door het *National Institute of Clinical Excellence (NICE)*⁹ in het Verenigd Koninkrijk ontwikkeld ten behoeve van evaluatie van digitale zorginnovaties. Op basis van de 21 beoordelingscriteria die het ESF hanteert, zijn 13 geaggregeerde categorieën gedefinieerd om de data uit de geïncludeerde studies in te verdelen. Een overzicht van de beoordelingscriteria gebruikt in dit rapport, en op basis van de welke ESF-criteria deze zijn vastgesteld, is te vinden in tabel 1. Studies die niet in voldoende detail een beschrijving gaven van welke specifieke beoordelingscriteria relevant waren voor de beoordeling en enkel een generieke omschrijving gaven, werden deze alsnog geëxcludeerd uit het overzicht voor beoordelingscriteria.

Tabel 1. Samenstelling van beoordelingscriteria op basis van ESF-criteria, zoals deze zijn gebruikt zijn binnen het literatuuronderzoek. DHT = Digital Health Technology.

Beoordelingscriterium	ESF-criteria
Safety	1: the DHT should comply with relevant safety and quality standards
Stakeholder involvement	2: incorporate intended user group acceptability in the design of the DHT 7: show processes for creating reliable health information 8: show that the DHT is credible with professionals
Technical development & validation	5: embed good data practices in the design of the DHT
DHT user	6: define the level of professional oversight
Equity & inequalities	4: consider health and care inequalities and bias mitigation
Privacy & GDPR	9: provide safeguarding assurances for DHTs where users are considered to be in vulnerable groups, or where peer-to-peer interaction is enabled
Environmental	3: consider environmental sustainability
Potential value & claims	10: describe the intended purpose and target population 11: describe the current pathway or system process 12: describe the proposed pathway or system process using the DHT 13: describe the expected health, cost and resource impacts compared with standard or current care or system processes
(Clinical) outcomes & study design	14: provide evidence of the DHT's effectiveness to support its claimed benefits

⁹ National Institute for Clinical Excellence (NICE). Evidence standards framework (ESF) for digital health technologies. Published December 10, 2018. Accessed April 20, 2023. <https://www.nice.org.uk/corporate/ecd7>

	15: show real-world evidence that the claimed benefits can be 16ategori in practice
Cost(-effectiveness)	17: provide a budget impact analysis 18: for DHTs with higher financial risk, provide a cost-effectiveness analysis
Post marketing surveillance (risks/harms)	16: the company and evaluator should agree a plan for measuring usage and changes in the DHT's performance over time
Implementation & scalability	19: ensure transparency about requirements for deployment 20: describe strategies for communication, consent and training processes to allow the DHT to be understood by end users 21: ensure appropriate scalability
Miscellaneous	-

Data-extractie werd uitgevoerd door één onderzoeker en gecontroleerd door een tweede onderzoeker. Na extractie vanuit de individuele rapporten werden de resultaten puntsgewijs samengevat in een beschrijvende lijst per beoordelingscriterium. Indien meerdere rapporten eenzelfde concept beschreven, werd dit slechts een keer meegenomen. Naast inhoudelijke informatie werden ook specifieke bronnen en richtlijnen geëxtraheerd (bijv. de Medical Device Regulation [MDR]) die per beoordelingscriterium werden aangehaald.

Aangezien een aangepaste vorm van het NICE ESF-raamwerk werd gebruikt als basis voor extractie van beoordelingscriteria, en dit raamwerk tevens de meest gedetailleerde informatie verschafte voor elk van de criteria, werden deze informatie separaat naast een samenvatting van de overige referenties weergegeven.

2.1.2 Beoordelingsprocedure

In overleg met ZIN werden regelgevende organisaties van vijf landen geselecteerd om de beoordelingsprocedures van digitale zorginnovaties in kaart te brengen. Dit waren:

- Oostenrijk: Austrian National Public Health Institute (Gesundheit Österreich GmbH, GoeG)
- België: Federaal kenniscentrum voor de gezondheidszorg (KCE)/mHealthBELGIUM
- Duitsland: Federal Institute for Drugs and Medical Devices (Bundesinstitut für Arzneimittel und Medizinprodukte, BfArM)
- Verenigd Koninkrijk: National Institute for Clinical Excellence (NICE)
- Verenigde Staten: Food & Drug Administration (FDA)

In februari 2023 werd op de websites van genoemde regelgevende instanties gezocht naar relevante informatie en verwijzingen naar gerelateerde documentatie. Relevante gegevens over de beoordelingsprocedures werden door één onderzoeker geëxtraheerd en gestructureerd samengevat in een tabel met de volgende kolommen:

- Identificatie: hoe komen digitale zorginnovaties de beoordelingsprocedure binnen?

- Selectie / randvoorwaarden: hoe worden digitale zorginnovaties geselecteerd en welke randvoorwaarden worden gesteld?
- Evaluatie: welke beoordelingscriteria worden gehanteerd bij de evaluatie?
- Besluitvorming: hoe verloopt de besluitvorming t.a.v. vergoeding en/of implementatie, en wie zijn daarbij betrokken?

Hierbij dient te worden opgemerkt dat de beoordelingscriteria geëxtraheerd onder 2.1.1 de inhoudelijke details weergeven die een rol spelen bij het onderdeel *Evaluatie* in het beoordelingsproces.

KCE (België) bleek na de start van deze academische werkplaats, op 13 januari 2023, een rapport te hebben uitgebracht getiteld '*Evaluation of digital medical technologies*'.¹⁰ Aangezien in dit rapport een internationale vergelijking werd gemaakt, is deze ook als basis gebruikt voor de beschrijving van de procedures in Duitsland en Oostenrijk.

2.2 Kwalitatief onderzoek

Het doel van het kwalitatieve onderzoek binnen deze academische werkplaats was om het onderwerp verder uit te diepen onder de diverse stakeholders. Daartoe werden allereerst focusgroepen georganiseerd. Vervolgens werd een vragenlijst uitgezet met als doel de bevindingen uit de focusgroepen voor te leggen aan een bredere groep relevante stakeholders.

In overleg met het Zorginstituut is ervoor gekozen om binnen het kwalitatieve onderzoek de nadruk te leggen op de effectiviteitsbeoordeling en niet op de volledige breedte van het beoordelingsproces met alle bijbehorende criteria.

2.2.1 Focusgroepen

Potentiële deelnemers werden mondeling of per e-mail benaderd voor deelname aan de focusgroepen. Zij werden doelgericht geselecteerd binnen de netwerken van de onderzoekers en het Zorginstituut, vanwege hun kennis en ervaringen m.b.t. het evalueren van digitale zorginnovaties vanuit een wetenschappelijk perspectief (onderzoeker, methodoloog, richtlijnontwikkelaar) of vanuit hun rol bij het beoordelingsproces (zorgverzekeraars, het Zorginstituut). Er zijn uiteindelijk drie homogene focusgroepen samengesteld: twee met deelnemers vanuit een wetenschappelijk perspectief (zeven deelnemers op 3 oktober en vier deelnemers op 20 december 2023) en één met vijf deelnemers vanuit het perspectief van beoordelende instanties (op 24 november 2023). Namens het onderzoeksteam waren drie personen aanwezig, waarvan één de rol van gespreksleider had (KJ).

De focusgroepbijeenkomsten vonden plaats in het UMC Utrecht en duurden twee uur. Na een korte kennismaking volgde een introductie van het onderwerp door de gespreksleider, waarbij de resultaten van het verkennende literatuuronderzoek kort gepresenteerd werden. De groepsdiscussie die daarop volgde, was semigestructureerd aan de hand van gespreksonderwerpen die vooraf waren opgesteld

¹⁰ Evaluation of Digital Medical Technologies | KCE. Published January 13, 2023. Accessed March 24, 2023. <https://kce.fgov.be/en/evaluation-of-digital-medical-technologies>

door de onderzoekers (zie 'Bijlage 1. Details focusgroepen: deelnemers & topic lijst'). Waar nodig stelde de gespreksleider aanvullende vragen en lette erop dat iedereen aan het woord kwam. Eén van de andere aanwezige onderzoekers maakte aantekeningen en bewaakte de tijd. Met goedkeuring van de deelnemers werd een geluidsopname van de groepsdiscussie gemaakt.

Met behulp van de geluidsopnames schreef één van de onderzoekers (PH) een uitgebreide samenvatting van elke groepsdiscussie. Deze samenvattingen werden vervolgens geanalyseerd om de belangrijkste thema's te identificeren. De codering werd gedaan door één onderzoeker (PH) en gecontroleerd door een tweede (KJ).

De bevindingen uit de focusgroepen werden gebruikt bij het opstellen van de vragenlijst (zie 2.2.2) en zijn verwerkt in het uiteindelijke logic model.

2.2.2 Vragenlijst

De vragenlijst werd uitgezet onder relevante stakeholders, bij voorkeur met affiniteit met digitale zorg, te weten: innovatoren/ontwikkelaars, (richtlijn)methodologen, wetenschappelijk onderzoekers, ZIN, NFU, zorgverzekeraars, klinici, patiënten, notified bodies en de overheid.

Via de volgende partijen en netwerken werd de vragenlijst verspreid:

- Health Innovations Netherlands (HI-NL)
- The Healthcare Innovation Center (THINC.)
- Zorginstituut Nederland (ZIN)
- Federatie medisch specialisten (FMS)
- Het Nederlands Huisartsen Genootschap (NHG)
- Platform Transformatie Digitale en Hybride Zorg (Digizo.nu) (vertegenwoordiging IZA partners)
- Patiëntenfederatie Nederland
- Persoonlijk 18ategor onderzoekers

De vragenlijst bestond uit zowel meerkeuzevragen als open vragen. Allereerst was er een vraag om te achterhalen in welke categorie stakeholder een respondent viel. Daarna volgde inhoudelijke vragen over het proces van evalueren van de effectiviteit van digitale zorginnovaties. De vragenlijst werd zo opgebouwd dat de diverse stakeholders vragen beantwoordden die voor hen relevant waren.

Bij elke vraag kreeg de respondent allereerst kort achtergrondinformatie omtrent het thema en de reflectie van de bevindingen uit de focusgroepen, waarna gevraagd werd om een reactie hierop en eventuele aanvullingen.

Beoogde respondenten werden uitgenodigd om via een weblink deel te nemen aan de vragenlijst. Zij konden de vragen vervolgens anoniem beantwoorden. Na het versturen van de initiële uitnodiging volgden twee herinneringen: na een week en drie dagen voor de deadline. De survey sloot na 2,5 week. Beoogde respondenten werden uitgenodigd om de weblink te verspreiden onder collega's met kennis of ervaring op het gebied van evaluatie, implementatie en/of opschaling van digitale zorginnovaties.

De software die gebruikt werd om de data te verzamelen, was MS Forms. Resultaten werden beschrijvend gepresenteerd. De resultaten op de open vragen werden door twee onderzoekers (onafhankelijk van elkaar) thematisch geanalyseerd.

De opzet van de vragenlijst staat in 'Bijlage 2 Opzet vragenlijst'.

2.3 Logic model

Als basis voor het logic model werden de resultaten uit het verkennende literatuuronderzoek en het huidige SWP-beoordelingsproces gecombineerd. Voor het huidige SWP-beoordelingsproces werd gekozen voor een versimpelde vorm, die niet alle details van het gehele beoordelingsproces weergeeft, maar wel de belangrijkste procesmatige stappen bij identificatie, selectie, beoordeling, en besluitvorming. Dit werd afgestemd met het Zorginstituut.

De belangrijkste onderwerpen en inhoudelijke bevindingen die uit de focusgroepen en/of uit het vragenlijstonderzoek naar voren komen, werden in het logic model verwerkt.

3. Resultaten

3.1 Verkennend literatuuronderzoek

De initiële zoekopdracht naar literatuur via referenties uit de *Agenderingsnotitie pakketbeheer en digitale zorg* resulteerde in een dataset van 190 referenties. Hieruit zijn na selectie 13 referenties geïnccludeerd voor data-extractie.

Daarnaast werden 114 unieke organisaties geïnccludeerd vanuit de HTA-databases. Na initiële selectie bleven hier 35 van over. Omdat bij enkele organisaties meerdere documenten geïdentificeerd werden die potentiële beoordelingscriteria omtrent digitale zorginnovaties omschreven, betrof het totaal aantal geïnccludeerde rapporten 71. Na het screenen van de volledige tekst werden hiervan uiteindelijk 11 geïnccludeerd voor data-extractie.

Tijdens het beoordelen van artikelen en rapporten werd in sommige gevallen verwezen naar externe literatuur die niet in onze initiële zoekstrategie naar voren was gekomen. Daarnaast werd naar aanleiding van een overleg met vertegenwoordigers van het Kwaliteitsconsortium nog enkele organisaties benoemd die handmatig zijn gezocht, en indien relevant toegevoegd aan de set voor data-extractie. In totaal werden om deze redenen nog eens 14 referenties toegevoegd, waarmee een totaal van 38 referenties zijn gebruikt voor data-extractie van de beoordelingscriteria.

Een flowchart van het selectieproces is te vinden in Bijlage 3 ('Flowchart verkennend literatuuronderzoek').

3.1.1 Beoordelingscriteria

Een beknopt overzicht van de subcriteria behorende bij elk van de beoordelingscriteria beschreven in de geïnccludeerde referenties is terug te vinden in tabel 2. Daarnaast is een gedetailleerder overzicht te vinden in een separate rapportage (*AWP Digitale Zorg Overzicht beoordelingscriteria literatuuronderzoek*).

Tabel 2. Beknopt overzicht van beoordelings- en subcriteria geïdentificeerd in het verkennend literatuuronderzoek. DHT = Digital Health Technology

Beoordelingscriterium	Subcriteria
Safety	<ul style="list-style-type: none"> • Clinical safety • Service safety • User (Risks or side-effects) safety • Risk of misuse • Risk of unhealthy behavior • Adverse events assessment and monitoring • Responsible person / entity for handling incidences • Material safety (e.g. physical, allergies) • Device safety
Stakeholder involvement	<ul style="list-style-type: none"> • User involvement in design (e.g. textual readability, graphic design) • User testing • User experiences

	<ul style="list-style-type: none"> • User involvement in implementation • Multi-stakeholder consortium (e.g. primary, secondary, tertiary care, public health) • Don't provide concrete goals for implementation of a specific DHT • Content experts with proven expertise • Comprehensive stakeholder engagement plan • Identify all stakeholder groups • Engage with stakeholders consistently and appropriately over the development and implementation phase • Develop and review the content, structure, interface and flow of the intervention • Identify the best digital platforms for the target population • Identify and address any aspects of the intervention that may unintentionally increase inequity and digital exclusion • Discuss and ensure that users understand who the intervention is for, which behavior it is trying to change, its aims, any possible harms, the time needed to establish behavior change and how frequently users are likely to interact with the intervention.
Technical development & validation	<ul style="list-style-type: none"> • Sources of clinical information up to date • Describe training and validation datasets (which ones, sample size, labelling, ground truth, how data was collected, diversity, use of synthetic data, choice of thresholds) • Performance testing (load test, stress test, spike test, etc.) • Intuitive and user-friendly templates • Reliability and maintainability of hardware (sensors) and software: (e.g. failure rates, measurement error rates, hardware risks, recurrent bugs, security bugs in the software) • * Use of existing literature to support technical validity
DHT user	<ul style="list-style-type: none"> • General user information provision (e.g. product name, product definition, price and any subscriptions or in-app purchases, sources of funding, assessment, author credits, contact details) • Training (who, by whom, how long, what materials, multiple languages)
Equity & inequalities	<ul style="list-style-type: none"> • Link to ethical concepts • Autonomy • Accessibility • Understandability / literacy of users • Ease of installation and use • Multiple languages • Cultural background • All (potential) end-users should be included • Equitable distribution of costs and health effects in cost-effectiveness analysis • Bandwidth / storage demand • Usability for people with sensory impairments • Internet / phone access • Low-income
Privacy & GDPR	<ul style="list-style-type: none"> • Privacy statement & general terms and conditions • Security (IT) statement • Tracking permission (e.g. cookies) • Data storage: why, where, how, how long, who has access, who oversees? • Communication of user data rights • Communication of analyses on user data • Anti-virus, network security and communication protocols (i.e. for email, chat and video)

	<ul style="list-style-type: none"> • Authentication procedure • Communication of security breach • Financial disclosures • User anonymization • Re-use of data
Environmental	<ul style="list-style-type: none"> • None found
Potential value & claims	<ul style="list-style-type: none"> • Description of unmet problem the claim attempts to solve • Claim vs empirical evaluation method • Domain (end-users) vs study population • Risk and side-effects • Risk/benefit trade-off (pilot study) • Position in the care pathway (add-on, triage, replacement) • Change in premises, information systems, or care processes • Change in leadership and organization (e.g. users, support) • Describe what the product impacts (e.g. clinical outcomes, costs (e.g. budget impact), system processes) • Describe how product results in impact • Comparator can be DHT or non-DHT (e.g. care as usual) • Workability when time investment is required • Use of existing literature to support assumptions (e.g. systematic reviews, original clinical research, professional society guidelines, secondary data analysis, clinical trials)
(Clinical) outcomes & study design	<ul style="list-style-type: none"> • Impact analysis (societal perspective) • Concept evaluation plan based on clinical association, technical performance, and clinical validation (risk class dependent) • Patient individual perceptions (in-depth interviews, focus groups) • Patient overall perceptions (survey) • Diagnostic or prognostic accuracy (cross-sectional research) • Clinical outcomes (RCT, pragmatic trials, naturalistic experiments, (retrospective) cohort studies, case-control, single arm trials, intra-individual studies, case series) > Phase dependent, product dependent (i.e. novel, or similar to existing products) • Clinical outcomes alternative designs: Trials of Intervention Principles method (TIP), Multiphase Optimization Strategy Trials (MAST), Sequential multiple assignment randomized trial (SMART), Continuous evaluation of evolving behavioral intervention technologies (CEEBIT), Intervention Informatica (I2), Real time technology assessment (Real time TA) and Comparative Effectiveness Research (CER) • Quality of life measurement using disease specific questionnaires (map on health utility scores) • Quality of life using EQ-5D • Multi-dimensional outcomes • Outcome matrix • Quantitative and qualitative evidence (i.e. mixed methods) • Barriers and facilitators (e.g. self-management, usability, accessibility, reliability, privacy) • Patient reported outcome measures (PROMS) • Impact on patient-provider relationship • Mortality, morbidity, quality of life • Duration and severity of disease • Accuracy measures (e.g. sensitivity, specificity, PPV, NPV) • Consequences of false positives/negatives

	<ul style="list-style-type: none"> • Cyclic updating of evidence • Sample size • Follow-up • Real-world evidence (not vignette studies) • Contextual usage • Study quality (i.e. risk of bias)
Cost(-effectiveness)	<ul style="list-style-type: none"> • Costs / benefits analysis through early HTA • Cost-effectiveness, cost-utility, cost-benefit, cost-consequence analysis, multi-criterion decision analysis (MCDA) (depends on research question) • Uncertainty incorporation in analyses • Impact matrix on accessibility, organization, quality and safety of care, and costs, per stakeholder • Initial start-up costs (e.g. buildings, facilities, training) • Maintenance costs • Software updating costs • Rapid, iterative analyses using updated empirical evidence • Micro costing (if pricing information is unavailable) • Direct and indirect costs • Pricing of the product (e.g. value-based pricing) • Digital and non-digital comparator • Place in clinical pathway (i.e. add-on, replacement, triage) • (rapid) Evolution and changing performance • User time and experience • Consider uptake rate for mean cost per user • Societal perspective • Non-health benefits • Cost-sequence combined with impact inventory
Post marketing surveillance (risks/harms)	<ul style="list-style-type: none"> • Monitor safety and functioning (by producer) • Risk management plan • Monitoring of clinical data and end-user experiences (by independent organization) • Technical stability criteria • Monitor uptime • Monitor crashes • FAQ continuous updating • Security updates • Clinical practice variation
Implementation & scalability	<ul style="list-style-type: none"> • Improvement of implementation or use (in-depth interviews, focus groups, expert opinion) • Expansion of use through wider application (i.e. transferability) • Expansion of use in international setting • Incorporation within I or provide dedicated interface • Interoperability (e.g. export data to other digital sources, print) • Update product based on user feedback • Introduction plan for healthcare service provider • Leadership / champion / superuser for the product at a healthcare facility • Sharing best-practices • Adjust clinical practice, organization, policy when new evidence becomes available • Technical support (i.e. availability and access) • Uptake is influenced by completeness, accuracy, relevance and comprehension of the content

	<ul style="list-style-type: none"> • Incentives (e.g. financial) • Evidence-based behavioral change (e.g. goals and planning, feedback and monitoring, and social support) • Multi-stakeholder collaboration to share costs • Describe distribution channels, pricing, procurements, and route of reimbursement
Miscellaneous	<ul style="list-style-type: none"> • Evidence available on all assessment criteria for the product at the time of writing • Free DHTs should be kept out of the reimbursement system • Invest in quality of existing services rather than develop new ones • Prevent implementation of DHTs sponsored by institutes that adversely affect health (e.g. tobacco companies)

In het algemeen kan geconstateerd worden dat alle beoordelingscriteria uit het ESF-raamwerk in meerdere of mindere mate verder werden toegelicht in tenminste een van de geïncorporeerde referenties. Een uitzondering was het criterium duurzaamheid (*environmental sustainability*), dat enkel bij het ESF-raamwerk werd genoemd en bij geen van de andere rapporten of raamwerken.

Bij elk van de 12 overige beoordelingscriteria konden meerdere subcriteria worden onderscheiden die concretere invulling geven aan het bovenliggende criterium. Zo werd bijvoorbeeld bij het criterium gelijkheid (*equity*) verschillende subcriteria genoemd waarop een digitale zorginnovatie (in de ontwikkelfase) zou moeten worden beoordeeld, zoals autonomie (v.d. gebruiker), toegankelijkheid, in hoeverre het aansluit bij de geletterdheid van de gebruiker, en beschikbaarheid van documentatie in meerdere talen. Daarnaast werden softwarematige randvoorwaarden gesteld, zoals de beschikbaarheid van voldoende opslag, bandbreedte, en beschikbaarheid van mobiel (internet) signaal in alle regio's, zodat geen groepen worden uitgesloten van gebruik.

Niet alle criteria zullen van toepassing zijn op de volledige breedte van digitale zorginnovaties. Zo zullen digitale zorginnovaties die beogen verandering teweeg te brengen in het zorgsysteem, niet worden beoordeeld aan de hand van dezelfde criteria als digitale zorginnovaties die direct invloed hebben op diagnostiek of behandeling van individuele patiënten/cliënten. Dit wordt door NICE in hun ESF-raamwerk expliciet gemaakt via het *tier* systeem, waarin gedifferentieerd wordt tussen de volgende categorieën:⁷

- Tier A: Digitale zorginnovaties die beogen kosten en personeel inzet te besparen (geen directe impact op gezondheidsuitkomsten)
- Tier B: Digitale zorginnovaties die beogen burgers en patiënten zelfredzaam te laten zijn
- Tier C: Digitale zorginnovaties die beogen diagnose te stellen en behandelingskeuzes te leiden voor medische aandoeningen

Daarnaast zullen niet alle subcriteria in even grote mate van toepassing zijn voor de Nederlandse context. Zo is vrijwel in heel Nederland dekking van het mobiel netwerk, waardoor exclusie van specifieke individuen in afgelegen gebieden niet aan de orde is. Het kwalitatieve onderdeel van deze academische werkplaats zal dienen om te onderzoeken welke van deze (sub)criteriën in het bijzonder van belang zijn voor de Nederlandse context.

Enkel bij de criteria *Potential value / claims, (Clinical) Outcomes & study design* en *Cost(-effectiveness)*, en *Considerations for implementation and scalability* 25atego concrete onderzoeksopzetten genoemd. Bij het subcriterium *Clinical Outcomes* werd over de verschillende rapporten en raamwerken een breed scala aan onderzoeksopzetten genoemd, waaronder RCTs, pragmatische trials, naturalistische trials, (retrospectieve) cohortstudies, patiënt-controle studies, single arm trials, intra-individuele studies en case series. Hierbij werd de noodzaak van een bepaalde onderzoeksopzet gekoppeld aan de fase van ontwikkeling waarin een digitale zorginnovatie zich bevindt.

Bij de overige (sub)criteria werden niet expliciet onderzoeksopzetten genoemd. Dit valt deels te verklaren doordat enkele van de subcriteria gaan over het rapporteren van eigenschappen van de digitale zorginnovatie of procedures. Hierbij kan bijvoorbeeld gedacht worden aan het afgeven van *privacy* en *security statements* of beschrijven hoe wordt omgegaan met de opslag van (patiënt)gegevens.

Ten slotte werden er in de geïnccludeerde referenties ook enkele onderwerpen en overwegingen genoemd die niet onder de vooraf gedefinieerde criteria van het ESF-raamwerk vielen, zoals het advies om te voorkomen dat digitale zorginnovaties worden geïmplementeerd die door een industrie worden gesponsord welke de gezondheid negatief beïnvloedt (bijv. de tabaksindustrie). Deze en andere overwegingen staan onder *Miscellaneous* in tabel 2 en het bestand *AWP Digitale zorg Overzicht beoordelingscriteria literatuuronderzoek*.

3.1.2 Beoordelingsprocedure

Een overzicht van stadia van de beoordelingsprocedure bij elk van de vijf geïnccludeerde regelgevende internationale organisaties is te vinden in onderstaande tabel 3. In Bijlage 4 'Beschrijving beoordelingsprocedure per instantie en land' staat een narratieve uitwerking van de procedures per organisatie.

In het algemeen kan worden geconcludeerd dat er tussen organisaties grote verschillen zitten in de focus, details en complexiteit van de verschillende stappen van de beoordelingsprocedure. Waar bijvoorbeeld de FDA een uitgebreide (voor)selectie heeft met verscheidene stappen, voorwaarden, regelingen en procedures, zijn deze stappen een stuk minder uitgekristalliseerd bij bijvoorbeeld de GoeG.

Aangezien de organisaties elk ook net andere verantwoordelijkheden hebben en zich in een eigen uniek zorgstelsel bevinden, was het categoriseren van hun taken en verantwoordelijkheden in de stadia van de beoordelingsprocedure een complexe exercitie. Het onderdeel *Besluitvorming* wordt bij NICE bijvoorbeeld vrij summier beschreven (een uiteenzetting van de evaluatie van een digitale zorginnovatie middels een *MedTech innovation briefing* (MIB) die wordt aangeboden aan de National Health Service (NHS)), maar de besluitvorming zal mogelijkterwijs in meer detail terug te vinden zijn bij documenten die de procedures binnen de NHS beschrijven. Uitdieping van procedures in gelieerde organisaties ligt buiten de scope van dit project en is derhalve niet verder onderzocht.

Tabel 3. Overzicht van de beoordelingsprocedures voor digitale zorginnovaties in vijf landen.

Land; organisatie	Identificatie	Selectie / randvoorwaarden	Evaluatie	Besluitvorming (vergoeding/implementatie)
Austria; Austrian National Public Health Institute (Gesundheit Österreich GmbH, GoeG) <i>*framework is under development, currently just theoretical</i>	Directly from manufacturers; but DMTs developed in collaboration with healthcare insurances (sickness funds) would also be considered	<ul style="list-style-type: none"> DMTs (CE marked) 26 categorized as such and be available in German. DMTs with a therapeutic purpose highlighted as a priority. Diagnostics and primary prevention excluded. Specific focus on chronic illnesses. All types of DMTs (both DMTs for self-use by patient or those to be used by the patient in consultation or collaboration with healthcare professional). 	Initially necessary requirements <ul style="list-style-type: none"> Technology specific Regulatory Evidence No economic requirements, no ethical requirements Two additional frameworks recommended for evaluation: <ul style="list-style-type: none"> Digi-HTA framework (Finland); domains: <ul style="list-style-type: none"> Safety Effectiveness Cost Usability and accessibility Data security and protection HTA module for mobile health applications (MMA) developed by Cambridge University (theoretical framework applied to Australian context) 	Pathway for price setting has not yet been decided

<p>Belgium; KCE/mHealthBELGIUM</p>	<p>Manufacturers can voluntarily register their mobile apps</p>	<p>Only apps that have completed levels M1 (CE marking and GDPR) and M2 (interoperability and connectivity) can start the integration process in the M3 level (reimbursement). Criteria for interoperability and connectivity:</p> <ol style="list-style-type: none"> 1. Categorization depending on the sensitivity of the data 2. identification of the person in need of care 3. authentication of the user 4. verification of the therapeutic relationship 5. interoperability: exchange of structured personal data takes place via open standards 6. GDPR compliance declaration 	<p>Manufacturers must submit a dossier showing</p> <ul style="list-style-type: none"> • the clinical and/or socio-economic value of their app: the evidence (whether real-world data or a clinical study; various study designs) must show that app brings clear value to the specific care process • the wider benefit to society or savings outside of the healthcare setting in which the cost is generated. <p>Specific working group (composed of independent and experienced experts, representatives of the relevant healthcare providers, insurers, patients, and representatives of employers' and employees' organisations) assesses the mobile app and formulates advice based on:</p> <ul style="list-style-type: none"> • clinical evidence • feasibility • added value • budgetary impact 	<p>Reimbursement can be temporary or structural and can be implemented via the following existing mechanisms:</p> <ul style="list-style-type: none"> • <i>nomenclature</i> (list of reimbursed healthcare provisions delivered by healthcare providers). • a "<i>convention</i>", on tract between the NIHDI (National Institute for Health and Disability Insurance) and healthcare providers or healthcare delivery institutions <p>one cannot exclude, in theory, certain DMTs to be reimbursed via the existing procedure for inscription of medical devices on a <i>positive (generic or nominal) list</i>.</p> <p>Mobile health apps are currently financed by other means: via hospital innovation budget, by the federated entities (if app supports certain policies), or "additional advantages" offered by sickness funds</p>
--	---	--	---	--

<p>Germany; Federal Institute for Drugs and Medical Devices (BfArM)</p>	<p>Manufacturers can voluntarily register their mobile apps by submitting their application for the DiGA Fast Track Process</p>	<ol style="list-style-type: none"> 1. Low-risk medical device (CE class 1 and 2a) 2. The DMT used as support to <ol style="list-style-type: none"> a. the identification, monitoring, treatment or alleviation of disease b. the identification, treatment or alleviation or compensation of injuries or disabilities 3. Its main function should be based on digital technologies 4. Its main function should not merely consist in the collection of data. 5. It should be used autonomously by the patient or by the patient together with the healthcare professional. <p>Primary prevention, telemedicine, and telemonitoring not included.</p>	<p>Manufacturer must include proof of the following:</p> <ol style="list-style-type: none"> 1. Safety and suitability for use (confirmed by CE mark) 2. Interoperability 3. Further quality requirements <ul style="list-style-type: none"> • Robustness against incidents and user errors • Consumer protection • Ease of use • Support for healthcare providers • Quality of medical content conform accepted medical standards • Patient safety 4. Data protection requirements and information security according to the state of the art 5. Evidence of “positive healthcare effects” <ul style="list-style-type: none"> • either a medical benefit: morbidity, mortality, or quality-of-life: • or a patient-relevant improvement of structure and processes in health care: aimed at supporting the health behaviour of patients or integrating the processes between patients and healthcare providers <p>quantitative comparative study required</p>	<p>GKV-SV (Spitzenverband der Gesetzlichen Krankenkassen – Head Association of the Statutory Health Insurances) is responsible for the price negotiations for the pricing of the DMT. For the first 12 months the price is freely set by the manufacturer. Only after this period the price is negotiated with the GKV-SV. If no agreement is reached in the negotiations, an arbitration board sets the reimbursement price.</p>
---	---	--	--	---

<p>UK; NICE</p>	<p>NICE selects digital health technologies that are registered at the NHS Innovation Service (formerly Health Tech Connect). Alongside the Innovation Service is the NICE Medical Technologies Evaluation Programme (MTEP).</p>	<ol style="list-style-type: none"> 1. The technology must be suitable for development under one of our guidance programmes, and not currently in development. 2. It must be a new technology, or an innovative modification of an existing technology, which will have benefits to the patient or the NHS when judged against comparators. 3. The technology must have CE marks (or equivalent regulatory approval). The technology must also be available to the NHS or have plans to be launched in the NHS. <p>A DHT that is part of the Medical Technologies Evaluation Programme is likely to be selected if it meets the eligibility criteria and:</p> <ul style="list-style-type: none"> • it offers substantial benefits to patients, or the health and care system compared with current practice • the benefits are plausible, clearly described, easily understood, and supported by evidence • developing guidance would mean faster and more consistent adoption of the technology 	<p>Developers of DHTs should use the following frameworks:</p> <ul style="list-style-type: none"> • The Digital Technology Assessment Criteria (DTAC) • The Evidence Standards Framework (ESF) for digital health technologies (not mandatory) <p>NICE advice is produced by a faster process that does not include a recommendation on NHS use. Instead, NICE advice (such as MedTech innovation briefings; MIBs) includes</p> <ul style="list-style-type: none"> • a description of a regulated medical device or in vitro diagnostics • its place in the care pathway • a critical appraisal of the available evidence and some expert opinion. 	<p>Whilst national funding (through the Innovation and Technology Payment (ITP)) is no longer provided, the NHS Accelerated Access Collaborative is encouraging NHS trusts to sustain these innovations through local funding agreements.</p> <p>Recommendations provided by the Accelerated Access Collaborative (AAC):</p> <ul style="list-style-type: none"> • To be recommended for NHS funding we think DHTs must be authorized by MHRA, pass DTAC and be recommended by NICE. • Geographical variation is common, as funding, and funding decisions, are still mostly made at a local or regional level by local commissioning bodies (formerly Clinical Commissioning Groups (CCGs), now Integrated Care Systems (ICS)
-----------------	--	---	---	---

<p>USA; FDA</p>	<p>Four steps to bring a device to the market:</p> <ul style="list-style-type: none"> • Step One: Classify Your Device and Understand Applicable Regulatory Controls • Step Two: Select and Prepare the Correct Premarket Submission • Step Three: Send your Premarket Submission to the FDA and Interact with FDA Staff During Review • Step Four: Comply with Applicable Regulatory Controls Including the Establishment Registration and Device Listing 	<p>Medical devices (including software) determining whether a product is regulated by the FDA as a medical device</p> <ol style="list-style-type: none"> 1. Determine if product meets definition of a medical device 2. Determine if an appropriate product classification exists for product 	<p>Owners or operators of establishments that are involved in the production and distribution of medical devices intended for use in the U.S. are required to register annually with the FDA.</p> <p>The requirements for a specific DHT depend on the class it is in.</p> <ul style="list-style-type: none"> • Class 1: General Controls (low-risk devices) General controls consider the following factors: <ul style="list-style-type: none"> ○ Good Manufacturing Practices ○ Standards and Reporting Adverse Events to FDA ○ Register their establishments and list the medical devices they market with FDA ○ General recordkeeping requirements ○ Are in accordance with labeling regulations and cannot be adulterated or misbranded. • Class 2: General Controls With Special Controls, in addition to class 1 controls: <ul style="list-style-type: none"> ○ Labeling requirements (information that must be included on a product label) ○ Device specific mandatory performance standards ○ Device specific testing requirements • Class 3: General Controls and Premarket Approval (devices support or sustain life, are implanted in the body, or have the potential for unreasonable risk of illness or injury) -> premarket approval; manufacturer must prove that device is safe and effective. <p><u>Submission types</u></p> <ul style="list-style-type: none"> • Premarket notification; indicates that the Class 2 medical device is similar to others on the market. • Premarket approval; for Class 3 devices, must include data from all nonclinical studies and clinical studies 	<p>Not applicable</p>
-----------------	--	--	---	-----------------------

			<ul style="list-style-type: none"> • De novo classification request; a pathway to classify novel medical devices for which general controls alone, or general and special controls, provide reasonable assurance of safety and effectiveness for the intended use, but for which there is no legally marketed predicate device • Humanitarian device exemption; intended to benefit patients with rare diseases or conditions 	
--	--	--	---	--

3.2 Kwalitatief onderzoek

3.2.1 Focusgroepen

De onderwerpen die besproken zijn in de focusgroepen zijn in te delen in de volgende thema's: criterium 'Plegen te bieden', claims, risico-inschatting t.a.v. effectiviteitsvraagstuk (SWP), passend bewijs, onderzoeksopzetten, en de rol van innovatoren/leveranciers. Hieronder worden deze thema's toegelicht.

Criterium 'Plegen te bieden'

Bij digitale zorginnovaties kan het gaan om innovaties die taakverschuiving mogelijk maken, zowel tussen beroepsgroepen, als van zorgverlener naar bijvoorbeeld een app. Naast vragen t.a.v. verantwoordelijkheid, rijst ook de vraag of dan nog kan worden voldaan aan het *plegen te bieden criterium*. De deelnemers merkten op dat niet altijd goed begrepen wordt wat met dit criterium bedoeld wordt en dat het gaat om technieken die een *beroepsgroep* tot hun behandelarsenaal rekent. De beroepsgroep bepaalt, niet het Zorginstituut. Zorgverzekeraars zijn in eerste instantie verantwoordelijk om zorg in te kopen ten laste van de Zvw of Wlz die voldoet aan dit wettelijke criterium, als ook aan het criterium 'Stand van de Wetenschap en Praktijk'. Hiervoor kijkt men vooral naar professionele (behandel)richtlijnen van de beroepsgroep. Echter richtlijnen zijn vaak verouderd en er kan veel tijd overheen gaan voordat een (digitale) innovatie in een richtlijn wordt opgenomen. Deelnemers merkten op dat zorgverzekeraars mogelijk digitale zorginnovaties (te) snel laten afvallen, omdat deze nog niet (volledig) verankerd zijn in richtlijnen en om die reden niet zouden voldoen aan het criterium 'plegen te bieden'.

“Probleem is de volgorde: als je bij eerste een kruisje zet, dan wordt naar de rest niet gekeken. Daar zit een verbeterpunt, dat je misschien zelfs wel gewoon zegt dat het een zwaarder weegt dan het ander in het licht van waar we in de maatschappij voor staan.”

Claims

Van alle zorg die volgens het 'plegen te bieden' criterium tot het behandelarsenaal van een zorgprofessional behoort, is vervolgens wettelijk vereist dat het duidelijk moet zijn dat deze zorg voldoende bewezen effectief is conform het criterium SWP. Uit de discussies tijdens de focusgroepen kwam naar voren dat veel van de claims bij digitale zorginnovaties zich echter niet (alleen) op effectiviteit richten, maar (ook) bijvoorbeeld op doelmatigheid, organisatie of continuïteit van zorg, adherentie bij cliënt/patiënt en gebruiksvriendelijkheid. Deze aanvullende claims komen deels terug in de overige pakketcriteria (kosteneffectiviteit, noodzakelijkheid en uitvoerbaarheid). Hier wordt in sommige gevallen aanvullend naar gekeken en alleen in geval van een positieve SWP-beoordeling (gezondheidswinst; het criterium SWP is een 'knock-out' criterium). Binnen het huidige kader geldt dat, op grond van de wettelijke criteria, nieuwe interventies ten minste even effectief moeten zijn als de huidige standaard- of gebruikelijke zorg. Zoals één van de deelnemers verwoordde: “er is dan geen ruimte voor een interventie die 5% minder (maar nog wel voldoende) effectief is en wel 90% goedkoper plus heel arbeidsbesparend”. Sommige deelnemers verwachten dat in het licht van de maatschappelijke opgaven een dergelijke afweging in de toekomst mogelijk gemaakt zal moeten worden.

Duurzaamheid en arbeidsinzet kwamen in de discussies naar voren als belangrijke maatschappelijke criteria om te wegen, maar deze komen in het huidige beoordelingskader nog niet expliciet terug. Opgemerkt werd dat digitale zorginnovaties ook kunnen leiden tot een hogere belasting van zorgpersoneel, ook al zouden ze bedoeld zijn om deze te verlagen.

In de focusgroepen kwam ook terug dat de huidige maatschappelijke opgave erom vraagt dat er breder gekeken wordt dan enkel naar effectiviteit. In zekere zin pleit dit voor een pragmatische omgang met het criterium SWP en mogelijk een zwaardere weging van andere (pakket)criteria.

Risico inschatten t.a.v. effectiviteitsvraagstuk

De vraag in welke gevallen een SWP-beoordeling nodig is, kwam als belangrijk thema naar voren in alle drie de focusgroepen. Dit is niet altijd zo zwart wit. Als voorbeeld werd het verplaatsen van fysieke zorg naar online zorg genoemd, waarbij (een deel van) de behandeling via een videoverbinding plaatsvindt of waarbij zelfs geen zorgverlener meer aan te pas komt. Het behandelen van een cliënt of patiënt door een psycholoog via een videoverbinding is wezenlijk anders dan een GGZ-module die een volledig autonome behandeling biedt voor depressie. Is dat laatste een technische variant of moet het gezien worden als een nieuwe interventie, omdat de factor menselijkheid (i.e. behandelrelatie) ontbreekt? Soms is daarnaast nog onduidelijk of de inzet van een digitale zorginnovatie nog onder aansprakelijkheid van de zorgverlener valt, en daarmee aan het criterium 'plegen te bieden' wordt voldaan. En nog een stap verder zijn op artificiële intelligentie (AI) gebaseerde innovaties. De deelnemers stelden dat verschillende typen digitale zorginnovaties onderscheiden kunnen worden met ieder een ander risicoprofiel.

Of een SWP-beoordeling nodig is, hangt niet alleen af van het type digitale zorginnovatie, het hangt volgens de focusgroepdeelnemers ook af van hoe groot het verschil is t.o.v. huidige, niet-digitale interventie en op welk onderdeel van het zorgpad het verschilt. Een digitale innovatie toevoegen aan een zorgpad geeft een ander risico dan een bestaande (fysieke) interventie helemaal vervangen. Ook is het belangrijk om na te gaan of de indicatie waarvoor de innovatie wordt toegepast anders of breder is dan die van de oorspronkelijke innovatie.

Een kenmerk van digitale zorginnovaties is verder dat ze voortdurend worden aangepast. Ook al is zo'n aanpassing niet gericht op (het verbeteren van) de effectiviteit, de effectiviteit zou er wel door beïnvloed kunnen worden (zie ook paragraaf *Onderzoeksopzetten*).

Passend onderzoek vragenlijst

De passend onderzoek vragenlijst die het Zorginstituut gebruikt om een inschatting te maken van welke studieopzet nodig en haalbaar is gegeven de context, werd met de deelnemers van de focusgroepen besproken.

In alle drie de focusgroepen werd genoemd dat de eisen aan het bewijs voor laag-risico digitale zorginnovaties anders zouden kunnen zijn dan voor hoog-risico digitale zorginnovaties (bijvoorbeeld geen RCT nodig, maar bewijs van lagere kwaliteit of minder bewijs kan acceptabel zijn).

Ook als er maar een bepaald onderdeel van een interventie of zorgpad verandert, dan zou een RCT niet in alle gevallen nodig zijn. Wanneer bijvoorbeeld van een combinatie diagnostiek en therapie het diagnostiekdeel vervangen zou worden door een digitale innovatie, dan zou, onder de aanname dat het therapiedeel niet veel verandert, enkel naar accuratesse gekeken kunnen worden. Verder zou in sommige gevallen de wetenschappelijke onderbouwing voor onderliggende principes of mechanismes kunnen volstaan.

“We hebben het over digitale zorginnovaties, maar het is zo breed. Gaat het om organisatie, dezelfde zorg in andere vorm geleverd, of zit er ook nog intelligentie in? Dat zijn totaal andere dingen, waarvoor misschien wel andere bewijslast nodig is. Maar dat begint bij interventie zelf.”

Deelnemers benadrukten de maatschappelijke opgave en transitie die gemaakt moet worden, en gaven aan dat in het kader hiervan de focus in eerste instantie kan liggen op laag risico-applicaties die geen verstrekkende gevolgen hebben.

Onderzoeksopzet

Tijdens de focusgroepen lag de nadruk op de effectiviteitsbeoordeling en niet op de volledige breedte van het beoordelingsproces. Een thema dat in alle focusgroepen terugkwam, is de continue doorontwikkeling van digitale applicaties, met verschillende versies tot gevolg. Hierdoor worden uitdagingen gezien bij het uitvoeren van RCTs, die over het algemeen een lange looptijd hebben. Vanuit het ontwikkelaarsperspectief kan men niet zo lang wachten (i.v.m. concurrentie), maar er zijn ook ethische bezwaren: in studieverband zou deelnemer versie 1 moeten blijven gebruiken, terwijl er al updates zijn. Er werd tijdens de focusgroepen in vrij algemene zin over updates gesproken, zonder expliciet onderscheid tussen technische- of veiligheidsupdates en updates die eigenschappen toevoegen waardoor de functie of indicatie van een toepassing beïnvloed worden. Wel werd aangehaald dat voor op AI gebaseerde innovaties de gevolgen van een update groter zouden kunnen zijn dan voor een statisch algoritme. Overigens vond één van de deelnemers dat het argument dat continue doorontwikkeling van een digitale zorginnovatie een RCT onmogelijk maakt, een oneigenlijk argument is. Deze deelnemer stelde dat een gezondheidsclaim op elk moment in de tijd onderzocht kan worden door ofwel te accepteren dat situatie anders is dan in maand 1, of de situatie bevroren voor duur van de studie.

“Bij digitale zorg wil je weten voor wie het werkt, welke onderdelen wel, welke niet; dat zijn procesgerichte studies. Dat het transformatie is naar een andere manier van zorgverlening en zorg implementeren, impliceert dat je op zoek bent naar een ander evaluatiekader. RCT's hebben een beperkte waarde daarin. Het is niet zo dat ze onnodig zijn, maar niet alléén RCTs.”

Tijdens de focusgroepen kwam enkele keren naar voren dat digitale zorginnovaties beschouwd moeten worden als complexe interventies, die die steeds verder ontwikkeld worden, lerende van de ervaring in de praktijk. Daarbij wil men niet alleen weten óf het werkt, maar is ook geïnteresseerd in wat werkt voor wie en wanneer (mechanismes ontrafelen). Het gaat daarbij om de toegevoegde

waarde van een digitale zorginnovatie binnen de specifieke context. Deelnemers benoemden de beperkte waarde van RCTs daarin en gaven aan dat er (ook) kwalitatieve onderzoeksmethoden nodig zijn (*mixed methods designs*), bijvoorbeeld in de vorm van een procesevaluatie.

Ten aanzien van de haalbaarheid van RCTs werd genoemd dat het uitvoeren van een RCT duur is en dat het voorstelbaar is dat er innovatoren zijn die dat helemaal niet kunnen betalen. Gesuggereerd werd om RCTs die nodig zijn (bij 'hoog risico' [zie eerder] en/of groot maatschappelijk belang) te vergoeden uit een speciaal fonds. Ook werd benoemd dat een RCT niet altijd haalbaar bij zeer lokale toepassingen, bijvoorbeeld in geval van een applicatie die in een ziekenhuis ontwikkeld wordt, waarbij het onderliggende model zelfs per afdeling kan verschillen).

Een van de deelnemers gaf een voorbeeld van een situatie waarin een RCT (nog) niet wenselijk is. Applicaties worden vaak vrij intuïtief ontwikkeld, waardoor de ontwikkelfase niet gedegen genoeg is doorlopen en bijvoorbeeld bruikbaarheid onvoldoende is aangetoond. Als dergelijke applicaties toch in een RCT geëvalueerd worden, dan is dat eigenlijk te vroeg en dit kan gevolgen hebben voor de effectiviteit die vervolgens gemeten wordt.

Monitoring werd genoemd als mogelijk alternatief in situaties waarin een RCT niet haalbaar is. De innovatie wordt dan al toegepast in de praktijk (en vergoed) en parallel daaraan worden data verzameld om effectiviteit (en andere relevante uitkomsten) te kunnen onderzoeken. Als blijkt dat de innovatie uiteindelijk niet effectief is, dan zou de vergoeding weer moeten worden teruggedraaid. Onvoldoende (kwaliteit van) registratie van data wordt als belangrijke barrière gezien voor monitoring. Dataverzameling loopt soms via leveranciers en niet bijvoorbeeld rechtstreeks via het elektronische patiëntendossier. Dit kan belemmerend werken door wederzijdse belangen en eisen, en ook door technische uitdagingen.

Verder werd benoemd dat er uitdagingen zijn om causale relaties aantonen o.b.v. *real-world evidence* en dat hier methodologisch onderzoek naar gedaan wordt.

“Passende zorg is het nieuwe buzz-woord. Passende zorg is eigenlijk effectieve zorg die waarde gestuurd is, want wat effectief is, betekent nog niet dat het voor jou waardevol is en ook zorg op de juiste plek zit. Je ziet veel digitale innovaties die daar wat aan proberen te doen: dat je substitutie van zorg krijgt, maar dan digitaal ondersteund. Daar ontbreekt nu eigenlijk een goed beoordelingskader/rechtskader, bij substitutie van bestaande verzekerde zorg door een nieuwe digitale variant.”

Rol innovatoren/leveranciers

Opgemerkt werd dat er een contrast is tussen de evaluatie van medicijnen waarbij grote farmaceutische bedrijven betrokken zijn met veel ervaring en geld versus de kleine startups in de wereld van (digitale) innovaties. Die laatste groep heeft over het algemeen beperktere financiële middelen en is minder ervaren in het opzetten en uitvoeren van klinische studies. Innovatoren/leveranciers moeten aan steeds meer criteria voldoen (wet- en regelgeving, Leidraad AI)

en certificering kost veel tijd en geld. Ook is lang niet altijd duidelijk wat er wel of niet onder medische software valt.

3.2.2 Vragenlijst

In totaal hadden 49 respondenten de vragenlijst ingevuld, waarvan 43 de vragenlijst volledig hadden ingevuld, en 6 partieel.

3.2.2.1 Resultaten gesloten vragen

De respondenten van de vragenlijst hadden een gevarieerde achtergrond (tabel 4). De grootste groep identificeerde zich als onderzoeker en/of (richtlijn)methodoloog. Andere affiliaties waren breed verdeeld over verschillende achtergronden. De groep overige affiliaties bestond onder meer uit consultants en (strategisch) adviseurs voor diverse organisaties.

Tabel 4. *Affiliaties deelnemers vragenlijstonderzoek*

Affiliatie	Aantal (%)
Onderzoeker / (richtlijn)methodoloog	22 (32%)
Beoordelende instantie	11 (16%)
Overig	9 (13%)
Ontwikkelaar / innovator	8 (12%)
Zorgprofessional	7 (10%)
Gebruiker digitale zorg	5 (7%)
Zorgprofessional vertegenwoordiger	4 (6%)
Evaluerende instantie	2 (3%)
Patiënt(vertegenwoordiger)	1 (1%)

Er waren vier dichotome gesloten vragen, die gingen over de thema's criterium 'Plegen te bieden', methoden voor bepalen technische variant, en de bruikbaarheid en compleetheid van de passend onderzoek vragenlijst. Een overzicht van de antwoorden is te vinden in tabel 5. Responses waren vergelijkbaar wanneer wetenschappers / richtlijnmethodologen met respondenten met andere affiliaties werden vergeleken (Bijlage 3. Kwantitatieve analyse vragenlijst: subgroep o.b.v. affiliatie).

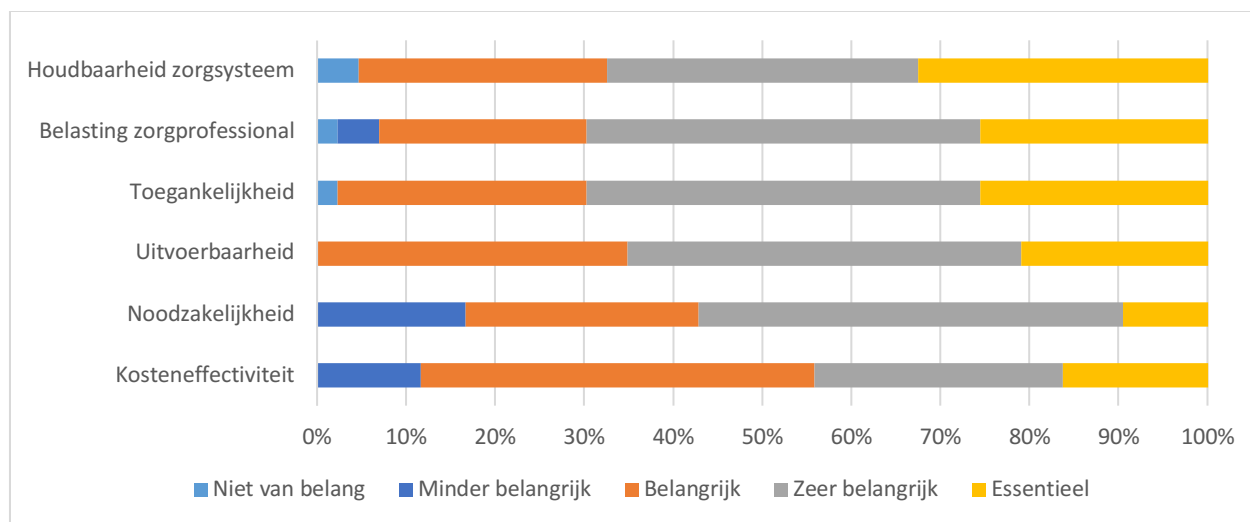
Tabel 5. *Reacties respondenten vragenlijstonderzoek op gesloten vragen*

	Aantal 'Ja' (%)
Komt u in de praktijk uitdagingen tegen omtrent het 'plegen te bieden' criterium in geval van digitale zorg?	29 (60%)
Bent u bekend met methoden om te bepalen of digitale zorg een technische variant of een nieuwe zorgvorm betreft?	23 (48%)
Zo ja, loopt u daarbij tegen uitdagingen aan?	12 (52%)
Is de passend onderzoek vragenlijst wat u betreft bruikbaar en geschikt voor digitale zorg?	27 (64%)
Ontbreken er vragen op de passend onderzoek vragenlijst of zijn er bepaalde vragen overbodig bij beoordeling van digitale zorg?	24 (62%)

Daarnaast werd gevraagd naar hoe zwaar enkele criteria zouden moeten wegen bij de beslissing of een digitale zorginnovatie moet worden geïmplementeerd. Er werd gevraagd naar reeds bestaande overige overwegingen (kosteneffectiviteit, noodzakelijkheid, uitvoerbaarheid), maar ook naar andere overwegingen die tijdens de focusgroepen werden genoemd (belasting zorgprofessional, houdbaarheid zorgsysteem, en toegankelijkheid van zorg).

Hoe respondenten aankeken tegen het belang van enkele overige overwegingen (naast de beoordeling van effectiviteit) staan weergegeven in Figuur 1. Het valt op dat de houdbaarheid van het zorgsysteem, belasting zorgprofessional, toegankelijkheid en uitvoerbaarheid, de hoogste proportie responses (65-70%) hebben wanneer wordt gekeken naar een zeer belangrijk of essentieel criterium. Noodzakelijkheid en kosteneffectiviteit werden minder vaak als zeer belangrijk of essentieel bestempeld, en beide door >10% van de respondenten als minder belangrijk gelabeld. Wanneer wetenschappers / richtlijnmethodologen werden vergeleken met respondenten met andere affiliaties, waren vergelijkbare patronen te zien (Bijlage 3. Kwantitatieve analyse vragenlijst: subgroep o.b.v. affiliatie)

Figuur 1. Belang van overige overwegingen bij beoordeling van digitale zorginnovaties



3.2.2.2 Resultaten open vragen

Er zijn 12 open vragen gesteld om gedetailleerde input te verkrijgen omtrent de verschillende aspecten van het evaluatieproces van digitale zorginnovaties.

Er zijn 6 thema's gedefinieerd, te weten: criterium 'Plegen te bieden', technische variant, typen digitale zorg, passend onderzoek vragenlijst, onderzoeksopzet, en overige overwegingen. Naast deze thema's zijn er enkele algemene opmerkingen gemaakt (bijv. in het kader van de gekozen terminologie), en verschillende bronnen met achtergrondinformatie aangereikt.

De uiteenzetting hieronder volgt deze onderwerpen, waarbij de 6 thema's genummerd zijn en middels deze nummering zijn gekoppeld aan de verschillende onderdelen van het logic model (zie paragraaf 3.3).

Het onderstaande is een synthese van de responses op de open vragen uit de vragenlijst. Een uitvoerigere uiteenzetting van de responses is te vinden in Bijlage 4.

1. 'Plegen te bieden' criterium

Het 'plegen te bieden' criterium kan een punt van discussie vormen wanneer bepaalde beroepsgroepen zorg erbij krijgen (overschot) of juist wanneer zij zorg moeten loslaten. Daarnaast wordt benoemd dat met name verplaatsing van zorg tussen de 1ste, 2e, en/of 3de lijn een uitdaging vormt. Een bijkomende uitdaging betreft dat beroepsgroepen veelal nog geen standpunt hebben ingenomen of een innovatie tot hun behandelarsenaal gerekend kan worden. Het innemen van een standpunt kan tevens gevolgen hebben voor verwachtingen van patiënten, die door digitale zorg die continu (thuis) monitort, verwachten dat zorgprofessionals ook onbeperkt bereikbaar dienen te zijn.

Er worden verscheidene concrete voorbeelden genoemd van digitale zorg waarbij het criterium 'Plegen te bieden' een punt van discussie vormde, zoals onder andere digitale groepstherapie, digitale triage, *telemonitoring*, en thuismonitoring door patiënten zelf (waar dit in een eerdere situatie door een zorgprofessional gebeurde). Zo werd in het kader van *telemonitoring* genoemd dat het onduidelijk was wie de hoofdbehandelaar, en wie er verantwoordelijk is voor het signaleren van een afwijkende waarde? Daarnaast wordt digitale triage genoemd, waarbij de verantwoordelijkheid verandert gedurende de verschillende stadia van triage (invullen van de tool, wachten in de digitale 'wachtkamer', afhandeling van het behandeladvies of plan).

“Er is onvoldoende comfort om “los te laten”, en te weinig ruimte om “erbij te krijgen” en dat speelt door in de verantwoordelijkheden”

2. Technische variant

De term 'technische variant' wordt veelal niet herkend door veldpartijen, waardoor slechts van een deel van de respondenten input kon worden verzameld m.b.t. dit criterium.

Respondenten benoemden dat er onduidelijkheid bestaat wanneer een innovatie een technische variant betreft. Er wordt aangegeven dat er behoefte is aan voorbeelden van innovaties zodat meer inzicht verkregen kan worden in wanneer er wel (of niet) sprake is van een technische variant. Daarnaast wordt aangegeven dat de vraag of een innovatie een technische variant betreft kan worden beoordeeld met het *indirectness* criterium binnen de GRADE-systematiek.¹¹ Echter ontbreekt hier een duidelijke instructie hoe deze concreet geëvalueerd dient te worden. De aanname dat een technische variëte vergelijkbare effectiviteit heeft als de huidige zorg, is een die lastig te beoordelen is, en vaak niet hard kan worden gemaakt.

¹¹ Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines: 8. Rating the quality of evidence--indirectness. J Clin Epidemiol. 2011;64(12):1303-1310. doi:10.1016/j.jclinepi.2011.04.014

3. Typen digitale zorg

De manier waarop digitale zorginnovaties worden geëvalueerd, en welk bewijs passend is, kan mogelijk variëren op basis van tot welk type een bepaald soort digitale zorg behoort. Echter zijn er verschillende assen waarop digitale zorginnovaties geïllustreerd kunnen worden. Dit werd eerder geïllustreerd aan de hand van de *tiers* zoals deze binnen NICE ESF worden gehanteerd¹² (zie literatuuronderzoek, paragraaf *Beoordelingscriteria*). Hier is verder uitvraag naar gedaan binnen het vragenlijstonderzoek.

Respondenten waren verdeeld over de mogelijkheid om onderscheid te maken in verschillende typen digitale zorg. Er werd meermaals aangegeven dat onderscheid maken niet wenselijk was. Digitale zorg moet in de breedste zin hetzelfde worden beoordeeld als traditionele zorg. Praktische, of financiële bezwaren werden door enkele respondenten niet als valide argument gezien om anders te handelen. Andere respondenten, gaven aan dat onderscheid maken wel wenselijk was. Zij wezen meermaals op de *risk/benefit* verhouding van de zorginnovatie als belangrijkste eigenschap, waarbij gekeken kan worden naar informatie verkregen uit bijv. CE-risicoklasse. Er zou dan ook onderscheid kunnen worden gemaakt op basis van risicoclassificatie van de digitale zorginnovatie. Wanneer innovaties een laag risico met zich meebrengen kan dit juist mandaat geven om een RCT uit te voeren.

Andere overwegingen die werden genoemd waren de plaats in het zorgpad (bijv. triage, diagnose, behandeling, of zorgproces evaluatie), complexiteit/eenvoud van de innovatie, de context / setting / doelgroep waarin de innovatie wordt ingezet. Deze zitten deels verdisconteerd in de eerdergenoemde *risk/benefit* verhouding.

Voorbeelden die worden genoemd van innovaties waarbij een trial wel wenselijk is: alle behandelingen, innovaties met hoog risico voor patiëntveiligheid, innovaties die bestaande zorg vervangen. Voorbeelden van typen digitale zorginnovaties waar een RCT niet wenselijk is: communicatiemiddelen, monitoring, of innovaties waarbij de behandelrelatie onderdeel uitmaakt van de innovatie.

4. Passend onderzoek vragenlijst

Respondenten vonden de passend onderzoek vragenlijst op enkele punten inhoudelijk niet geschikt voor digitale zorg. Zo wordt genoemd dat het pathofysiologisch werkingsmechanisme iets is wat bij veel digitale zorginnovaties niet van toepassing is, en de bewoording voor digitale zorginnovaties zou moeten worden aangepast. Daarnaast is de vragenlijst gericht op aandoeningen, waar digitale zorginnovaties vaak gericht zijn op de vorm waarin bepaalde zorg geboden wordt, of breed inzetbaar zijn over indicatiegebieden heen. Hierbij kan worden gedacht aan bepaalde vormen van *telemonitoring*, of apps die een breed pallet aan klachten of vitale parameters kunnen monitoren.

¹² National Institute for Clinical Excellence (NICE). Evidence standards framework (ESF) for digital health technologies. Published December 10, 2018. Accessed April 20, 2023. <https://www.nice.org.uk/corporate/ecd7>

Overstijgend wordt genoemd dat de huidige passend onderzoek vragenlijst de afgelopen jaren niet bij iedere stakeholder heeft geleid tot het gedrag wat nodig is om tot adoptie en opschaling van de innovatie te komen. Passend bewijs zal maar een deel van het bewijs leveren wat nodig is om implementatie van de innovatie te bewerkstelligen.

Veel respondenten waren van mening dat de overige overwegingen (zie paragraaf hieronder) ook mee dienden te worden genomen bij de passend onderzoek vragenlijst. Juist bij digitale zorginnovaties spelen deze criteria een belangrijke rol.

Tevens werd genoemd dat de vragenlijst sterk georiënteerd is op behandeling, niet (digitale) ondersteuning. Veel digitale zorginnovaties meten patiëntkarakteristieken, en zijn daarmee diagnostisch of prognostisch van aard. Deze innovaties passen minder goed in de vragen dan traditionele geneesmiddelen die therapeutisch van aard zijn.

Door de respondenten worden een aantal factoren genoemd die mogelijk onderdeel uit zouden kunnen maken van de passend onderzoek vragenlijst, zodat deze beter zou kunnen passen voor digitale zorginnovaties. De meest genoemde zijn vragen omtrent de context waarin de innovatie wordt ingezet, en de impact welke deze heeft op verstoring van het zorgproces.

Daarnaast worden met name factoren besproken die onder de overige overwegingen (zie volgende paragraaf) vallen, zoals kosten, gelijkheid, haalbaarheid, noodzakelijkheid, etc.

“Passende hybride zorg met inzet van digitale toepassingen is m.i. vaak een organisatorisch, veranderkundig of financieel vraagstuk naast een medisch inhoudelijk vraagstuk.”

5. Onderzoeksopzet

Verskillende onderzoeksopzetten kunnen worden overwogen om de effectiviteit van digitale zorginnovaties aan te tonen. In grote lijnen kan onderscheid worden gemaakt tussen experimentele en observationele studies, waarbij respectievelijk wel en niet door onderzoekers wordt ingegrepen in de dagelijkse klinische praktijk (bijv. door het geven van medicatie, een diagnostische test, of onderwerpen aan een digitale zorg app). Daarnaast wordt het begrip *real-world evidence* (RWE) steeds vaker genoemd. Dit betreft bewijs dat verzameld is als onderdeel van routinepatiëntenzorg.^{13,14,15}

¹³ FDA. Real-World Evidence. FDA. Published April 25, 2023. Accessed May 2, 2024.

<https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence>

¹⁴ Jansen MS, Dekkers OM, le Cessie S, et al. Real-World Evidence to Inform Regulatory Decision Making: A Scoping Review. Clin Pharmacol Ther. Published online February 23, 2024. doi:10.1002/cpt.3218

¹⁵ Dang A. Real-World Evidence: A Primer. Pharmaceutical Medicine. 2023;37(1):25. doi:10.1007/s40290-022-00456-6

In de internationale literatuur bestaat geen formele consensus over de mate waarin RWE observationeel of experimenteel van aard is. Echter, impliceert de term ‘routinezorg’ dat er geen interventie (door onderzoekers) plaatsvindt, en het daarmee per definitie observationeel van aard is. Daarnaast wordt RWE vaak gecontrasteerd met (traditionele) RCTs.¹⁴ Onder de respondenten werd tevens enkele keren opgemerkt dat RWE hetzelfde is (of dezelfde uitdagingen heeft) als observationeel onderzoek. Bij de vraag over het gebruik van data uit RWE bij de beoordeling van effectiviteit werd meermaals verwezen naar antwoorden op de voorgaande vraag over het gebruik van data uit observationeel onderzoek bij de beoordeling van effectiviteit.

Naar aanleiding hiervan is besloten om in het kader van deze rapportage de gegeven antwoorden omtrent belemmerende en bevorderende factoren, en overwegingen bij onderzoeksopzet van observationele en RWE studies samen te voegen. Zie voor verdere toelichting de begrippenlijst aan het begin van dit document.

Hieronder worden belemmerende en bevorderende genoemd voor deze onderzoeksopzetten. Daarnaast worden suggesties gedaan voor adaptaties van deze onderzoeksopzetten, om zo tegemoet te komen aan de belemmerende factoren die worden herkend.

“De RCT blijft de gouden standaard m.b.t. streven naar unbiased uitkomsten. Echter, deze rigor is in de meeste gevallen mosterd na de maaltijd (het is er al...) en weegt totaal niet op tegen kosten (en wie moet dit betalen?), de doorlooptijd en als belangrijkste argument: de externe validiteit van digitale innovaties.”

Randomized controlled trials (RCTs)

Belemmerende factoren

Als voornaamste uitdaging wordt genoemd dat veel digitale zorginnovaties technologisch verouderd zijn voordat een RCT kan worden afgerond. Er is binnen trials vaak geen mogelijkheid om gaandeweg de innovatie aan te passen of een software update te doen. Genoegen nemen met een kortere duur van de RCT (bijv. 3-6 maanden) kan dit mogelijk tegengaan, maar dan wordt voorbijgegaan aan het feit dat er vaak onvoldoende power om met zekerheid conclusies te trekken, evenals de langetermijneffecten van de innovatie.

Daarnaast kunnen digitale zorginnovaties vaak in verschillende contexten worden ingezet, zoals verschillende lijnen in het zorgsysteem, populaties, of bij verschillende ziektebeelden. Deze context bepaalt voor een groot deel de effectiviteit van een innovatie. Juist binnen een RCT wordt een homogene populatie gekozen, die onder sterk gecontroleerde omstandigheden de innovatie toebedeeld krijgen.

Overstijgend wordt benoemd dat RCTs tot een afname in motivatie van het betrokken personeel tot gevolg heeft, alsmede extra druk legt in de vorm van administratielast op de al overbelaste beroepsgroepen.

RCTs voor digitale zorginnovaties zijn mogelijk niet haalbaar omdat deze een groter aantal studieparticipanten nodig hebben, met name bij settings met een lage ziekteincidentie (bijv. bij thuismonitoring), en veel financiële middelen vergen van vaak kleine *start-up* bedrijven.

Daarnaast is randomisatie veelal niet mogelijk, omdat innovaties reeds beschikbaar zijn in de dagelijkse praktijk. Patiënten willen zich vervolgens niet willen laten randomiseren, omdat ze zelf willen bepalen of ze gebruik maken van de digitale zorginnovatie. Data van individuen die niet de zorginnovatie ontvangen (controlegroep) is dan nog steeds te verkrijgen, maar omdat deze uit routine zorg systemen dienen te worden gehaald, is er risico op *confounding by indication*.¹⁶

Het gebrek aan mogelijkheden tot includeren en randomiseren van individuen bemoeilijkt niet alleen de snelheid van inclusie van studiedeelnemers in een klinische studie, maar ook de inclusie van deelnemende centra.

“Ik heb alleen ervaring met een RCT voor telemonitoring versus conventionele opname. Uitkomst: telemonitoring is net zo veilig, goedkoper en geassocieerd met grotere tevredenheid van de patient. So far so good. Maar uitvoering van de RCT was zeer uitdagend en duurde veel te lang”

Bevorderende factoren

Op voorhand dienen keuzes verantwoord te worden, zodat duidelijk wordt waarom een RCT noodzakelijk is. Deze dienen te worden gecommuniceerd aan alle betrokken stakeholders, zoals studiedeelnemers en centra.

Samenwerking is de sleutel tot succes. De respondenten noemen regionale of landelijke multi-stakeholder samenwerkingen en co-creatie om de haalbaarheid van een RCT te vergroten. Een gezamenlijke bekostiging, of specifieke subsidies voor startups, worden genoemd als factoren die financiële beperkingen kunnen terugdringen.

Op organisatorisch niveau wordt genoemd dat RCTs als vereiste zouden kunnen worden gesteld, waarbij de digitale zorg pas publiek beschikbaar zou moeten kunnen worden gesteld, zodra deze effectief (of als technische variant van bestaande zorg) wordt bevonden.

Overwegingen bij onderzoeksopzet RCTs

Er worden enkele aanpassingen op de traditionele RCT genoemd die mogelijk kunnen fungeren als facilitator om deze onderzoeksopzet voor digitale zorginnovaties haalbaarder te maken. Deze worden in de wetenschappelijke literatuur beschreven onder *alternative trial designs*.¹⁷ Deze zijn in de basis nog steeds een experimenteel design (en behoeven derhalve bijvoorbeeld

¹⁶ Joseph KS, Mehrabadi A, Lisonkova S. Confounding by Indication and Related Concepts. *Curr Epidemiol Rep.* 2014;1(1):1-8. doi:10.1007/s40471-013-0004-y

¹⁷ Harvin JA, Zarzaur BL, Nirula R, King BT, Malhotra AK. Alternative clinical trial designs. *Trauma Surg Acute Care Open.* 2020;5(1):e000420. doi:10.1136/tsaco-2019-000420

nog steeds goedkeuring van een METC), maar beogen de uitvoerbaarheid en externe validiteit te vergroten.

Zo wordt de pragmatische trial genoemd als alternatief op de traditionele (verklarende) RCT. Dit type trial beoogt een meer heterogene (en daarmee representatievere) studiepopulatie te verkrijgen, hetgeen de externe validiteit van de resultaten vergroot.¹⁸ Om grote studies, met meerdere betrokken centra, gemakkelijker te betrekken en te randomiseren naar de innovatie of een controle arm zonder dat er contaminatie optreedt, wordt de *stepped-wedge* cluster gerandomiseerde trial genoemd.¹⁹ Hierbij wordt de innovatie gefaseerd uitgerold over verschillende clusters (bijv. centra). Dit heeft enkele voordelen, zoals het organisatorisch uitrollen van de interventie, vergelijking van de innovatie tussen clusters maar ook over de tijd, en de mogelijkheid om lessen die geleerd zijn in de eerste clusters toe te passen in toekomstige clusters.

Naast deze twee alternatieven op het traditionele RCT-ontwerp, worden tevens het *platform design*²⁰ en *non-inferiority design*²¹ genoemd. Deze laatste kan in het bijzonder goed werken wanneer de innovator dient aan te tonen dat effectiviteit vergelijkbaar is (*non-inferior*, en daarmee dus een technische variant) in vergelijking met de huidige zorg.

Er bestaat ook de mogelijkheid om te verkennen of *interrupted time-series design*²² (ook wel quasi-experimentele design genoemd) haalbaar / wenselijk zou zijn bij een bepaalde innovatie. Deze bevatten geen randomisatie en halveren daarmee het aantal benodigde deelnemers.

Door tijdens een trial gebruik te maken van interim analyses, kan gaandeweg inzicht worden verkregen in de effectiviteit, nog voordat de trial formeel is afgerond. Dit kan kansrijke (of kansarme) innovaties vroegtijdig aan het licht brengen.

Real-world evidence (RWE)

Belemmerende factoren

RWE wordt door een aantal respondenten op zichzelf niet gezien als voldoende bewijskracht om effectiviteit aan te tonen, een RCT zal altijd nodig zijn. Een andere respondent gaf aan dat een hoge bewijskracht (*high level of evidence*) binnen de GRADE methode theoretisch wel mogelijk is, maar dit praktisch vaak niet haalbaar is. Observationeel bewijs zal binnen GRADE altijd starten als *low certainty of evidence*, waarbij *upgraden* noodzakelijk is om meer

¹⁸ Loudon K, Treweek S, Sullivan F, Donnan P, Thorpe KE, Zwarenstein M. The PRECIS-2 tool: designing trials that are fit for purpose. *BMJ*. 2015;350:h2147. doi:10.1136/bmj.h2147 <https://www.precis-2.org/>

¹⁹ Hemming K, Haines TP, Chilton PJ, Girling AJ, Lilford RJ. The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting. *BMJ*. 2015;350:h391. doi:10.1136/bmj.h391

²⁰ Pitre T, Cheng S, Cusano E, et al. Methodology and design of platform trials: a meta-epidemiological study. *J Clin Epidemiol*. 2023;157:1-12. doi:10.1016/j.jclinepi.2023.02.010

²¹ Hahn S. Understanding noninferiority trials. *Korean J Pediatr*. 2012;55(11):403-407. doi:10.3345/kjp.2012.55.11.403

²² Bernal JL, Cummins S, Gasparrini A. Interrupted time series regression for the evaluation of public health interventions: a tutorial. *Int J Epidemiol*. 2017;46(1):348-355. doi:10.1093/ije/dyw098

vertrouwen te krijgen in de gevonden effectiviteit.²³ Een factor die daartoe binnen GRADE bekeken wordt is de grootte van het gevonden effect, echter omdat digitale zorg vaak een relatief kleine impact op effectiviteit heeft, zal dit nooit resulteren in een *upgrade* van de *certainty of evidence*.

Daarnaast wordt benoemd dat het gebruik van reeds bestaande data infrastructures een uitdaging vormt voor de bruikbaarheid van analyse van RWE. Verschillende dataplatforms sluiten vaak niet op elkaar aan, en het koppelen hiervan kan uitdagingen geven met betrekking tot privacy eisen en informed consent. Ook ontbreken er vaak een groot aantal metingen, of complete variabelen, in reeds verzamelde datasets, en zijn kwaliteit en interpreteerbaarheid daarmee beperkt. Daarnaast wordt door commerciële partijen veelal niet de moeite genomen om additionele data (bijv. gebruikerservaringen, veiligheid, tijdsinvesteringen) te registreren. Goed gereguleerd databeheer en eigenaarschap ontbreekt veelal.

RWE studies brengen veelal uitdagingen in de vorm van bias met zich mee die het gevolg zijn van de routinematige verzameling van deze data. Zo is er risico op selectiebias, bijvoorbeeld door een bepaalde subset aan patiënten de digitale zorg te geven die extra gemotiveerd zijn, hetgeen geen representatieve resultaten geeft voor de gemiddelde patiënt die in de dagelijkse klinische praktijk zal worden gezien.

De analyse van dergelijke datasets wordt tevens vaak bemoeilijkt door gebrek aan statistische expertise om met de verschillende vormen van bias op een adequate manier om te gaan. Betrekken van personen met expertise binnen statistische analyse van dergelijke data, zeker in een vroeg stadium, is aanbevolen. Zelfs wanneer de expertise aanwezig is, dient ervoor gewaakt te worden dat big RW-datasets zich vooral lenen voor het genereren van hypothesen, niet bij uitstek om gerichte effectiviteitsvragen te beantwoorden.

Bevorderende factoren

Een groot deel van de respondenten zag RWE als aanvulling of alternatief op een RCT m.b.t. de mogelijkheid om voldoende (indirect) overtuigend bewijs op te leveren om effectiviteit van een innovatie aan te tonen. Ondanks dat GRADE-beoordeling start op een lagere *level of evidence*, kan met een grootschalig, methodologisch correct uitgevoerde observationele studie, welke een sterk effect laat zien, alsnog een *high certainty of evidence* verkregen worden.

RWE biedt verscheidene kansen als het gaat om dataverzameling. Zo kunnen naast de klinische uitkomstenmaten ook andere niet-klinische maten worden meegenomen, zoals onderdelen genoemd onder 6. *Overige overwegingen*. Dataverzameling kan plaatsvinden over langere perioden, grotere studiepopulaties omvatten, en er is mogelijkheid de innovatie of het zorgpad gaandeweg bij te stellen en te evalueren wat de impact is.

²³ Granholm A, Alhazzani W, Møller MH. Use of the GRADE approach in systematic reviews and guidelines. *British Journal of Anaesthesia*. 2019;123(5):554-559. doi:10.1016/j.bja.2019.08.015

Observationele data uit RWE studies bevatten een heterogenere onderzoekspopulatie, hetgeen inzicht kan geven in het effect van de innovatie in een representatiever sample dan in een RCT. Daarnaast geeft het representatiever weer wat de compliance is m.b.t. het gebruik van de innovatie.

Na implementatie van de innovatie kan het reeds ingerichte netwerk van databases gebruikt worden om de innovatie over langere perioden te monitoren. Koppelingen van databases zijn uitdagend, maar met een toenemende datasolidariteit en data infrastructuur zoals die van de huisartsen (COH), bieden kansen voor monitoring van lange-termijn effecten.

“RCTs vinden veelal toch plaats in een gecontroleerde klinische setting. Digitale zorg kent juist ook een grotere meerwaarde buiten klinische parameters. Parameters die met RWE beter te staven zijn.”

Overwegingen bij onderzoeksopzet RWE-studies

De structuur van RWE-verzameling dient meer gestandaardiseerd en verplicht te worden. Innovatoren zouden een vooraf goedgekeurd protocol moeten opstellen voor monitoring, zodat meerdere partijen akkoord moeten geven op de validiteit en kwaliteit ervan.

RWE biedt kansen voor het continu monitoren van patiënten. Echter dient vooraf goed nagedacht te worden over de data infrastructuur, privacy en toestemmingsverklaringen, en noodzaak van monitoring. Deze moet redelijkerwijs in verhouding staan met de risico's die er aan de innovatie gekoppeld zitten.

Zorg ervoor dat RWE-data-analyse door een onafhankelijke partij wordt uitgevoerd, zodat keuzes gemaakt tijdens de analyse niet kunnen leiden tot een vertekening van het resultaat van de studie.

Overige onderzoeksopzetten

Kortcyclisch onderzoek, met mogelijkheden om snel actie-lerend aanpassingen te maken in de innovatie of de inzet, wordt meermaals door respondenten genoemd als een methode om efficiënt digitale zorginnovaties te evalueren in een RW-context. Daarnaast wordt aangegeven dat haalbaarheidsstudies (bijv. *early HTA*, of kwalitatief onderzoek onder stakeholders), voorafgaand aan een grootschaligere empirische studie, inzicht kunnen geven in de kans op een succesvolle studie-uitkomst.

6. Overige overwegingen

Bij kosteneffectiviteit werd genoemd dat het belangrijk is om budgetimpact, PROMS, productiviteitsverliezen en investeringskosten mee te nemen.

Haalbaarheid werd door een respondent opgesplitst in acceptatie, *usability*, integratie, reikwijdte, praktische haalbaarheid, en implementatie.

Bij noodzakelijkheid werd opgemerkt dat het helder moet zijn wie het noodzakelijk acht. Het perspectief kan bepalend zijn bij beantwoording van deze vraag.

Omtrent belasting van de zorgprofessional werd opgemerkt dat het breder dient te worden gedefinieerd, waarbij ook mantelzorgers en patiënten zouden kunnen worden geïncludeerd. De belasting kan af- maar ook toenemen door introductie van een digitale zorginnovatie, omdat deze mogelijk de drempel tot de zorgverlener kan verlagen. In het bijzonder dient er bij digitale zorginnovaties aandacht te zijn voor de schaalbaarheid.

Additionele overwegingen

Enkele respondenten benoemden het overstijgende maatschappelijke doel van beoordeling van digitale zorginnovaties. Veel innovaties gaan verloren wanneer een RCT wordt geëist. Verzekeraars zien digitale zorginnovaties als middel om het acute zorg infarct dicht te lopen, en hebben derhalve minder behoefte aan valide vergelijkend onderzoek. Waak dan ook voor afwaardering van valide onderzoek door steeds soepeler om te gaan met RWE. Enkel overlaten aan de markt is niet wenselijk; er dient een gestandaardiseerd proces te komen.

“Digitale transformatie gaat om het oplossen van (toekomstige en nijpende) maatschappelijke problemen. De trial is in mijn ogen slechts in beperkte mate een instrument die in die queeste iets toevoegt.”

Naast de reeds genoemde en uitgevraagde overwegingen zijn additionele overwegingen genoemd door de respondenten. Een respondent gaf aan dat naast bewijskracht omtrent effectiviteit zou dienen te worden verdisconteerd met impact op de andere overige overwegingen.

Duurzaamheid wordt vaak genoemd als een belangrijke overweging om mee te nemen bij beoordeling van digitale zorginnovaties.

Daarnaast worden patiëntvoorkeuren, -tevredenheid, en gebruiksvriendelijkheid genoemd, alsmede ook het werkplezier van de betrokken zorgprofessional. Ook wordt *equity* vaker genoemd als een belangrijke component (bijv. toegang voor personen met minder digitale vaardigheden).

Arbeidsinzet wordt meermaals door respondenten genoemd als een belangrijke additionele overweging, naast de algehele tijdsinvestering die zorgprofessionals kwijt zijn bij de zorg voor hun patiënten / cliënten.

Ook worden een aantal praktische overwegingen genoemd die mogelijk een plek zouden kunnen innemen bij de beoordeling. Zo wordt gesproken over de uitvoerbaarheid, toekomstbestendigheid, en mogelijkheden tot het doorvoeren van (software)updates. Dit is

gerelateerd aan betrouwbaarheid en kwaliteit van de innovatie, iets wat tevens vaker in de vragenlijst werd genoemd.

Bronnen & achtergrondinformatie

Door een aantal respondenten werden verschillende bronnen en andere documenten die achtergrondinformatie bevatten aangehaald. Een overzicht van deze bronnen is onderaan Bijlage 4 terug te vinden. Deze bevat onder andere referenties naar terminologie en initiatieven omtrent digitale zorg, onderzoeksmethoden, en reflecties op overwegingen die andere (inter)nationale instanties hebben gemaakt in het kader van beoordeling van digitale zorginnovaties.

3.3 Logic model

De bevindingen uit het literatuuronderzoek en het kwalitatieve onderzoek (focusgroepen en vragenlijstonderzoek) worden samengevat in het *logic model* weergegeven in Figuur 2.

De verschillende fases in het beoordelingsproces zoals beschreven in het literatuuronderzoek is als structuur over het reeds bestaande SWP-beoordelingsproces gelegd. Het SWP-beoordelingsproces is versimpeld weergegeven in verschillende blokken. Met de rode vierkanten is aangegeven voor welke aspecten van het huidige SWP-proces er in de focusgroepen en het vragenlijstonderzoek in het bijzonder aandacht was. De belangrijkste bevindingen met betrekking tot de verschillende thema's staan onderaan de figuur.

De bevindingen zijn in te delen onder drie fases van het beoordelingsproces:

- 1) Voorafgaand aan SWP-beoordeling
- 2) SWP-beoordeling
- 3) Maatschappelijke weging

Uit het kwalitatieve onderzoek kwam naar voren dat voorafgaand aan een daadwerkelijke SWP-beoordeling er onduidelijkheid bestaat over welke digitale zorginnovaties een (volledige) SWP-beoordeling moeten (of kunnen) doorlopen. Zo wordt het 'plegen te bieden' criterium niet altijd goed begrepen en/of is het onduidelijk of een digitale zorginnovatie eraan kan voldoen (en er dus een SWP-beoordeling kan volgen). Inschatten of er een effectiviteitsvraag is bij een digitale zorginnovatie kan uitdagend zijn: wanneer is een digitale zorginnovatie bijvoorbeeld een technische variant en wanneer moet het beschouwd worden als nieuwe interventie? En hoe om te gaan met de continue doorontwikkeling die kenmerkend is voor veel digitale zorginnovaties? Of een SWP-beoordeling nodig is, hangt samen met het verschil ten opzichte van de bestaande situatie (niet-digitale interventie of zorgpad).

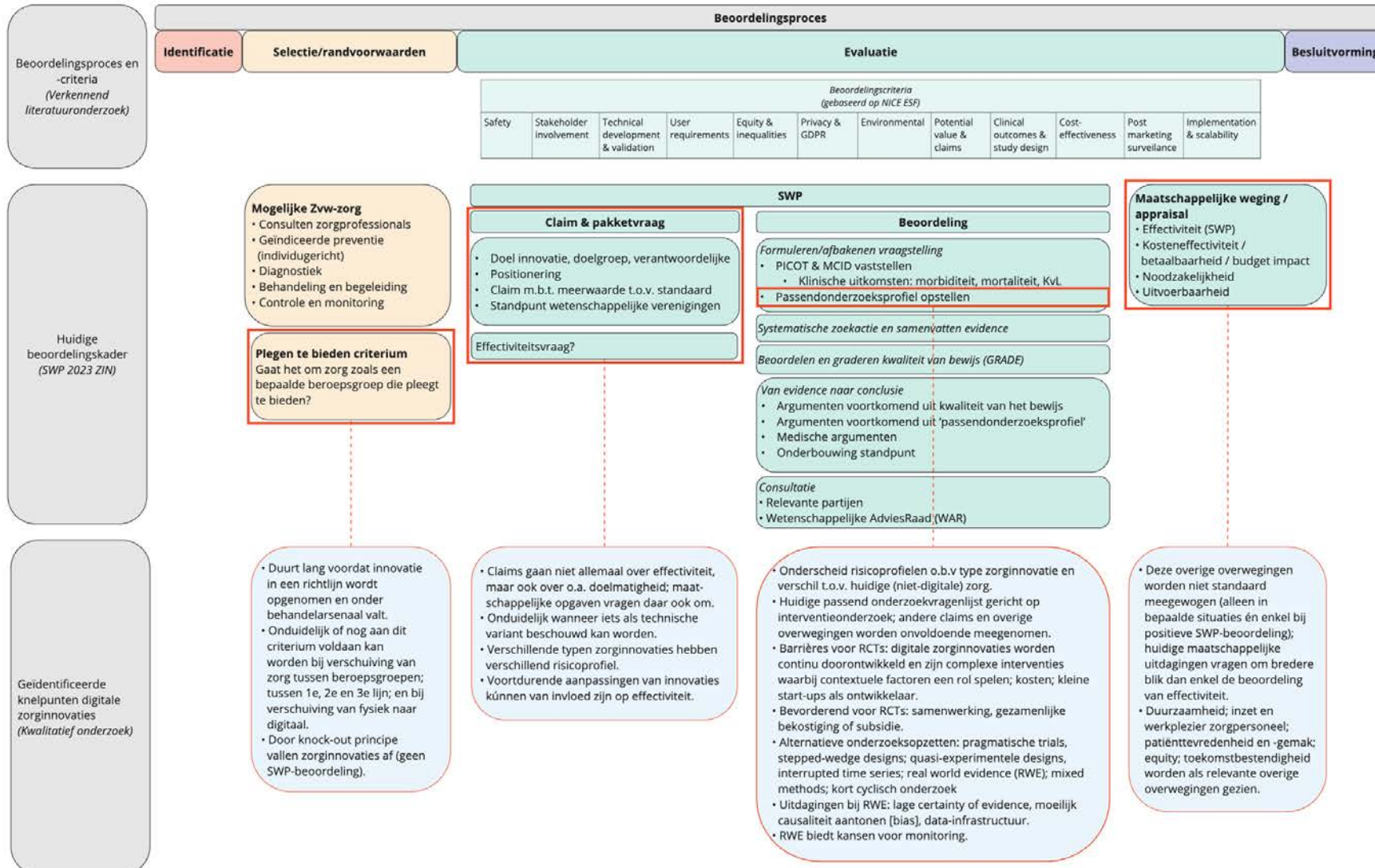
Daarnaast was één van de bevindingen in het kwalitatieve onderzoek dat het risico op nadelige gevolgen voor de gezondheid niet voor alle typen digitale zorginnovaties hetzelfde wordt ingeschat, denk aan *telemonitoring* versus inzet van *digital therapeutics*. In het verkennende literatuuronderzoek zagen we een dergelijk onderscheid tussen typen digitale zorginnovaties terug. Dit onderscheid is te maken op verschillende assen. Zo hanteert NICE drie categorieën gebaseerd op de gebruiker en het doel van de innovatie, en daarmee potentiële impact op gezondheidsuitkomsten. In Oostenrijk is deze

indeling overgenomen, maar het beschreven beoordelingskader beperkt zich daar tot therapeutische innovaties voor chronische ziekten. In Duitsland wordt in het beoordelingsproces onderscheid gemaakt tussen laag-risico en hoog-risico innovaties (o.b.v. CE-categorie). *Telemedicine* en *telemonitoring* worden als aparte categorie gezien, vanwege de significante input die zorgprofessionals hierbij hebben. Nictiz hanteert op vergelijkbare manier indeling op basis van gebruiker en plaats in het zorgproces, maar voegde hier de dimensie type technologie aan toe.

Met een onderscheid tussen typen digitale zorginnovaties en bijbehorende risicoprofielen zou vervolgens rekening gehouden kunnen worden tijdens de SWP-beoordeling door de eisen aan het bewijs af te stemmen op het risicoprofiel. Bij een laag-risicoprofiel zou bewijs van lagere kwaliteit of minder bewijs acceptabel zijn en een RCT niet per se nodig. Er is een scala aan alternatieve onderzoeksopzetten genoemd, variërend van een RCTs in aangepaste vorm tot observationele studies en gebruikmaken van *RWE*. Deze alternatieve onderzoeksopzetten bieden mogelijkheden om tegemoet te komen aan belemmerende factoren, zoals grootte van de benodigde studiepopulatie en organisatorische uitdagingen bij multi-center studies. Aangezien digitale zorginnovaties veelal geschikt zijn en gebruikt worden om geautomatiseerd, grote hoeveelheden data te verzamelen, worden kansen gezien voor gebruik van *RWE* bij de evaluatie ervan. Monitoring werd genoemd als mogelijk alternatief voor situaties waarin een RCT niet haalbaar is: de innovatie wordt al toegepast in de praktijk (en vergoed) en parallel daaraan worden data verzameld om effectiviteit (en andere relevante uitkomsten) te kunnen onderzoeken. Onvoldoende (kwaliteit van) registratie van data en uitdagingen om causale relaties aan te tonen worden als belangrijke barrières gezien voor monitoring op basis van *RWE*. Bij het verkennende literatuuronderzoek werden ook voorbeelden gezien van een rol voor alternatieve onderzoeksopzetten bij de evaluatie van digitale zorginnovaties. In Oostenrijk hangt het af van het type digitale zorginnovatie of een RCT nodig is (dit wordt van geval tot geval bekeken) en KCE (Federaal kenniscentrum voor de Gezondheidszorg in België) gebruikt expliciet de term *RWE* als gegevensbron. In Duitsland daarentegen is minimaal een kwantitatieve, vergelijkende studie vereist om effectiviteit aan te tonen.

Veel van de claims bij digitale zorginnovaties richten zich niet (alleen) op effectiviteit, maar bijvoorbeeld (ook) op doelmatigheid, organisatie of continuïteit van zorg. Het gaat dan over pakketbeheer in de breedte. De overige pakketcriteria (kosteneffectiviteit, noodzakelijkheid en uitvoerbaarheid) worden alleen in bepaalde gevallen beoordeeld en enkel als gezondheidswinst is aangetoond (positieve SWP-beoordeling). Dit 'knock-out principe' werd in het kwalitatieve onderzoek als belemmering genoemd. Duurzaamheid en arbeidsinzet worden gezien als criteria die voor de brede maatschappelijke opgave relevant zijn, maar deze zijn nog niet wettelijk verankerd bij het bepalen van verzekerde zorg. De socio-economische waarde speelt in België een rol bij het bepalen of een digitale zorginnovatie tijdelijk vergoed wordt (als de socio-economische waarde nog niet is aangetoond) of structureel (als de socio-economische waarde is aangetoond). Daarnaast zijn schaalbaarheid van de innovatie, duurzame inzet, arbeidsinzet, en informatievoorziening en ondersteuning aan innovatoren factoren die tijdens implementatie van de zorginnovatie nauwlettend gemonitord dienen te worden.

Figuur 2. Logic model beoordelingsproces SWP in het kader van digitale zorg



4. Discussie

Binnen deze academische werkplaats zijn diverse aspecten van de beoordeling van digitale zorginnovaties onderzocht. De voornaamste bevindingen, uit zowel literatuuronderzoek als uit focusgroepen en een vragenlijstonderzoek, zijn samengevat in een *logic model*.

Allereerst werden via een verkennend literatuuronderzoek de beoordelingscriteria ten behoeve van evaluatie van digitale zorginnovaties, zoals beschreven in wetenschappelijke en niet-wetenschappelijke literatuur, uiteengezet en de beoordelingsprocessen van verschillende (inter)nationale beoordelende instanties in kaart gebracht. De processen van alle beoordelende instanties zijn te 'mappen' op vier fases: identificatie, selectie, evaluatie en besluitvorming. De beoordelingscriteria beschrijven verschillende facetten van digitale zorginnovaties en bevatten vaak verschillende subcriteria die gebruikt kunnen worden om de criteria te operationaliseren. Effectiviteitsbeoordeling is slechts één van vele factoren waarnaar gekeken kan worden bij de beoordeling van digitale zorginnovaties.

Het tweede deel van deze academische werkplaats richtte zich specifiek op de effectiviteitsvraag binnen het beoordelingskader SWP en bestond uit empirische dataverzameling bij stakeholders in het veld van digitale zorginnovaties middels focusgroepen en een vragenlijstonderzoek.

Overstijgend bleek dat er onduidelijkheid is omtrent het beoordelingsproces SWP. Niet alle stakeholders begrijpen wanneer er sprake dient te zijn van een SWP-beoordeling (nl. enkel wanneer er een effectiviteitsvraag ligt) of wat het 'Plegen te bieden' criterium exact inhoudt. Stakeholders vinden het daarnaast uitdagend te beoordelen wanneer een innovatie een 'technische variant' betreft. Erkend en herkend werd dat er verandering noodzakelijk is, omdat de eis van gerandomiseerde studies voor het aantonen van effectiviteit voor elke vorm van digitale zorginnovatie, botst met de huidige uitdagingen omtrent belasting van zorgprofessionals en algehele druk op de zorg. Afhankelijk van het risicoprofiel van digitale zorginnovaties op basis van type zorginnovatie en het verschil ten opzichte van de huidige (niet-digitale) interventie of zorgpad zouden alternatieve onderzoeksopties ingezet kunnen worden om een robuust evidenceportfolio op te bouwen. Er lijkt een rol weggelegd voor monitoring op basis van RWE, omdat veel digitale zorginnovaties geschikt zijn om (geautomatiseerd) grote hoeveelheden data te verzamelen. Naast effectiviteit zijn claims van digitale zorginnovaties (ook) gericht op andere uitkomsten, zoals doelmatigheid, organisatie of continuïteit van zorg, wat brede is dan de beoordeling SWP. Duurzaamheid en arbeidsinzet werden genoemd als waardevolle aanvullende criteria die voor de brede maatschappelijke opgave relevant zijn.

De maatschappelijke urgentie voor transparantie en standaardisering van de evaluatie van digitale zorginnovaties blijkt ook uit andere initiatieven in Nederland die eveneens trachten het brede perspectief van digitale zorginnovaties te vangen. Het Digizo initiatief²⁴ is voortgevloeid uit een samenwerking van alle IZA-partijen en kijkt met een breed perspectief naar verschillende facetten van de evaluatie van digitale zorginnovaties, niet enkel focussend op effectiviteit. Hetzelfde geldt voor de

²⁴ Digizo. <https://digizo.nu/> Accessed February 27, 2024.

Leidraad AI²⁵, die breed naar het proces van ontwikkeling, evaluatie en implementatie kijkt, maar een focus heeft op digitale zorginnovaties in de vorm van voorspelmodellen gebaseerd op AI. Effectiviteit vormt bij beide initiatieven slechts een klein onderdeel van de beoordelingsstructuur. De uitkomsten van deze academische werkplaats bieden juist gedetailleerde overwegingen omtrent passend onderzoek en onderzoeksopzetten en vormen daarmee een aanvulling op deze andere initiatieven.

Implicaties / aanbevelingen

De bevindingen uit deze academische werkplaats geven richting voor mogelijke vervolgstappen. Er is allereerst behoefte aan duidelijkheid rondom het beoordelingsproces: welke digitale zorginnovaties dienen een (volledige) beoordeling SWP te doorlopen? Voorbeelden aan de hand van casuïstiek rondom het ‘plegen te bieden’ criterium en ‘technische variant’ zouden hieraan kunnen bijdragen.

Omdat niet alle digitale zorginnovaties eenzelfde impact op gezondheid hebben, zou met betrekking tot het benodigde evidenceportfolio gedacht kunnen worden aan stratificatie op basis van risicoprofielen, gerelateerd aan type innovatie. De indeling die NICE, CE-classificatie, DIGA, Nictiz, en andere organisaties hanteren zouden hiervoor als inspiratie kunnen dienen. Met een dergelijk onderscheid zou vervolgens rekening gehouden kunnen bij de evaluatie door de eisen aan het bewijs af te stemmen op het risicoprofiel. De huidige passend onderzoek vragenlijst kan worden gebruikt om een dergelijke afweging te maken. De weging van RWE (of andere alternatieve onderzoeksopzet) ten opzichte van RCT-data kan plaatsvinden binnen de reeds bestaande GRADE-systematiek.

Gezien het brede karakter van digitale zorginnovaties wordt echter gepleit voor het toekennen van meer aandacht en gewicht aan de overige overwegingen naast de beoordeling van effectiviteit. Dit past bij de maatschappelijke uitdagingen omtrent druk op de zorg en zorgprofessionals. Bij de implementatie dient rekening te worden gehouden met een breder pallet aan factoren, zoals schaalbaarheid van de innovatie, duurzame inzet, arbeidsinzet en aan welke informatievoorziening en ondersteuning innovatoren behoefte hebben.

Ten slotte

De bevindingen uit deze academische werkplaats kunnen als basis dienen voor toekomstige projecten in het kader van de evaluatie van digitale zorginnovaties. Verdere uitwerking van het kader kan zorgprofessionals, wetenschappers, beleidsmakers en het brede publiek transparantie en helderheid verschaffen en innovatoren helpen bij het verzamelen van passend bewijs voor een evidenceportfolio van hun innovatie.

²⁵ Van Smeden, M., Moons, K. G., Hooft, L., Chavannes, N. H., van Os, H. J., & Kant, I. (2023, October 19). Guideline for high-quality diagnostic and prognostic applications of AI in healthcare. <https://doi.org/10.17605/OSF.IO/TNRJZ>

Bijlagen

Bijlage 1. Details focusgroepen

Deelnemers

Focusgroep 3 oktober 2023

- Prof. dr. Patrick Bossuyt, Epidemiology & Data Science, Amsterdam UMC
- Prof. dr. Lisette van Gemert-Pijnen, eHealth & Wellbeing Research, Universiteit van Twente,
- Dr. Ilse Kant, portfoliomanager Digital Health, UMC Utrecht
- Dr. Kim Luijken, assistant professor, UMC Utrecht
- Dr. Tuur Leeuwenberg, assistant professor, UMC Utrecht
- Dr. Linda Dusseljee-Peute, director and principal educator eHealth Living & Learning Lab MI, Amsterdam UMC
- Dr. Sofie Willems, ANIOS Huisartsgeneeskunde, assistant professor NELL, LUMC

Focusgroep 24 november 2023

- Ron van Asselt, medisch adviseur Zorginstituut Nederland
- Prof. dr. Wolfgang Ebbers, bijzonder hoogleraar ICT and Strategic Innovation in the Public Sector, Erasmus University Rotterdam
- Dennis Japink, Zorgverzekeraars Nederland; Kenniscentrum Digitale Zorg
- Geert-Jan van Kemenade, medisch adviseur Zorginstituut Nederland
- Robin Toorneman, Senior Adviseur Zorginnovaties, Zorginstituut Nederland

Focusgroep 20 december 2023

- Dr. Anefloor van Enst, Eigenaar Medical Guidelines; Directeur Nederlandse Vereniging voor Dermatologie en Venereologie
- Carin Smit, arts-onderzoeker, Nederlandse Vereniging voor Dermatologie en Venereologie
- Dr. Jaap Trappenburg, THINC. Healthcare, UMC Utrecht
- Robin Toorneman, Senior Adviseur Zorginnovaties, Zorginstituut Nederland

Namens onderzoeksteam bij alle drie de focusgroepen:

- Dr. Pauline Heus, assistant professor, Cochrane Netherlands, UMC Utrecht
- Prof. dr. Lotty Hooft, hoogleraar Evidence Synthesis and Knowledge Translation, Cochrane Netherlands, UMC Utrecht
- Dr. Kevin Jenniskens, assistant professor, Cochrane Netherlands, UMC Utrecht

Topiclijst

1. Welkom en kennismaking (10 minuten)
2. Introductie (10 minuten)
 - a. Context, afkadering & doel AWP Digitale zorg
 - b. Resultaten verkennende literatuuronderzoek
3. Groepsdiscussie, inclusief pauze (90 minuten)
 - a. Praktische gang van zaken
 - b. Onderwerpen
 - Typen claims
 - Onderzoeksdesigns
 - Uitkomsten (van effectiviteit)
 - Passend bewijs profiel
 - GRADE-beoordeling
 - Overige overwegingen
4. Afsluiting (10 minuten)

Introductie

Deze vragenlijst gaat over de beoordeling van de effectiviteit van digitale zorg.

De vragenlijst is onderdeel van het project 'Van black-box naar white-box, beoordeling en evaluatie van innovaties' binnen de Academische Werkplaats Zorg en Beleid van Zorginstituut Nederland (ZIN) en de Nederlands Federatie van Universitaire Medische Centra (NFU).

Dit project richt zich op de evaluatie van digitale zorg. Digitale zorg is "de toepassing van zowel digitale informatie als communicatie om de gezondheid en gezondheidszorg te ondersteunen en/of te verbeteren." (naar de definitie van E-health van Nictiz)²⁶

Er kan grofweg onderscheid gemaakt worden in digitale toepassingen in het primaire (patiëntenzorg) proces, het secundaire proces (administratieve- en ondersteuningsprocessen) en op volksgezondheidsniveau (preventie, gezondheidsvoorlichting). We richten ons in deze survey op digitale zorg binnen het primaire zorg proces.

Eerder werd in het kader van deze Academische werkplaats een verkennend literatuuronderzoek uitgevoerd en zijn focusgroepen gehouden. Via deze vragenlijst toetsen we de eerdere bevindingen bij een bredere groep stakeholders.

Wij nodigen u uit om onderstaande vragen te beantwoorden.

De vragenlijst bestaat uit zowel meerkeuzevragen als open vragen. Indien u onvoldoende achtergrondkennis heeft om een vraag te beantwoorden, dan kunt u deze vraag overslaan. Uw antwoorden zijn niet herleidbaar naar u als persoon en het invullen kost u 20 minuten.

Door onderstaande vragenlijst in te vullen gaat u akkoord met de (geanonimiseerde) verwerking van uw antwoorden.

²⁶ Nictiz, E-health, wat is dat? 2019. <https://nictiz.nl/publicaties/e-health-wat-is-dat/>

Algemeen

Welk van onderstaande omschrijvingen is op u van toepassing? Meerdere antwoordopties zijn mogelijk. [meerkeuzevraag: meerdere antwoordopties mogelijk]

- a. Ontwikkelaar / innovator van (digitale) zorg
- b. Onderzoeker / (richtlijn)methodoloog
- c. Werkzaam bij een instantie die (digitale) zorg beoordeelt (zoals ZIN of zorgverzekeraar)
- d. Werkzaam bij instantie die (digitale) zorg evalueert (zoals VILANS)
- e. Vertegenwoordiger van gebruikers van (digitale) zorg
- f. Zorgprofessional
- g. Vertegenwoordiging beroeps- of branchevereniging van zorgprofessionals
- h. Anders, namelijk [Vak voor vrije tekst]

Achtergrondinformatie

Alle vergoede zorg (zowel nieuwe als bestaande zorg) binnen de Zvw en de Wlz moet voldoen aan de wettelijke criteria ‘Plegen te bieden’ en ‘Stand van de wetenschap en praktijk’ (SWP). Deze criteria zijn dus ook van toepassing op digitale zorg. Hieronder de definities:

Plegen te bieden

Zorg die ‘pleegt te worden geboden’ betreft zorg die een beroepsgroep rekent tot het aanvaarde arsenaal van zorg, en die geleverd wordt op een wijze die de beroepsgroep beschouwt als professioneel juist.²⁷ Het gaat in de context van deze survey om de vraag of van de digitale zorg die geëvalueerd wordt, redelijkerwijs verondersteld kan worden dat deze onder het aanvaarde zorgarsenaal van een beroepsgroep valt.

Vaak kan aan de hand van richtlijnen en standaarden worden vastgesteld of sprake is van zorg die de beroepsgroep ‘pleegt te bieden’.

Stand wetenschap en praktijk (SWP)

Aan de hand van het beoordelingskader SWP wordt bepaald of zorg op groepsniveau in voldoende mate **bewezen effectief** is conform de principes van *evidence-based medicine*. Hierbij wordt nagegaan of een interventie (diagnostiek, behandeling) resulteert in een relevante (meer)waarde voor de doelgroep ten opzichte van de standaardbehandeling. Er moet sprake zijn van netto winst op uitkomstmaten die voor de doelgroep/patiënt relevant zijn.

Beoordelingsstadia

Het beoordelingsproces van effectiviteit kent grofweg drie stadia. Deze survey zal worden doorlopen aan de hand van deze stadia. Deze betreffen:

- Selectie voor beoordeling
- Verzamelen, beoordelen en graderen van wetenschappelijk bewijs
- Overige overwegingen (‘from evidence to decision’)

²⁷ <https://www.zorginstituutnederland.nl/publicaties/rapport/2008/11/17/betekenis-en-beoordeling-criterium-plegen-te-bieden>

Selectie voor beoordeling

- 1) Digitale zorg roept in de praktijk soms vragen op omtrent het ‘plegen te bieden’ criterium. Digitale zorg kan namelijk leiden tot een verschuiving in verantwoordelijkheden of tot onduidelijkheid over wie de verantwoordelijkheid draagt. Bij nieuwe digitale zorg is ook niet altijd helder of de beroepsgroep de zorg al rekent tot het geaccepteerde behandelarsenaal.

Komt u in de praktijk uitdagingen tegen omtrent het ‘plegen te bieden’ criterium in geval van digitale zorg? Zo ja, kunt u die situatie(s) beschrijven?

- a. Nee
 - b. Ja, te weten... [Toelichting]
- 2) Soms is bij digitale zorg sprake van bestaande, bewezen effectieve zorg, die in een andere vorm dan de oorspronkelijke wordt aangeboden. Denk bijvoorbeeld aan teleconsulten i.p.v. fysieke consulten of een voedingsdagboekje in een app i.p.v. geschreven formulieren. Dit heet een ‘technische variant’. Onder voorwaarden kan het effectiviteitsbewijs dan worden overgenomen en hoeft de effectiviteit niet opnieuw te worden aangetoond. Partijen vinden het in de praktijk vaak moeilijk om in te schatten of er sprake is van een technische variant of van ‘nieuwe zorg’.

Bent u bekend met methoden om te bepalen of digitale zorg een technische variant of een nieuwe zorgvorm betreft? Zo ja, loopt u daarbij tegen uitdagingen aan? Licht uw antwoord toe

- a. Nee
- b. Ja, maar ik ervaar geen uitdagingen
- c. Ja, en ervaar uitdagingen namelijk [Toelichting]

Verzamelen, beoordelen en graderen van wetenschappelijk bewijs

- 3) Bij het aantonen van de effectiviteit van (digitale) zorg wordt het beoordelingskader SWP toegepast. Daarvoor wordt de kwaliteit van het wetenschappelijke bewijs beoordeeld en gewogen conform de principes van *evidence-based medicine* (EBM), en geplaatst in de medische context/klinische praktijk. De kwaliteit van bewijs wordt gewogen samen met eventuele andere argumenten.

Goed opgezette en uitgevoerde randomized controlled trials (RCTs) geven de hoogste kwaliteit van bewijs voor het beantwoorden van een effectiviteitsvraag. Tegelijkertijd kennen RCTs ook uitdagingen. Ontwikkelingen van digitale zorg gaan snel en RCTs vergen vaak tijd en investering, waardoor ze niet in alle situaties wenselijk dan wel haalbaar zijn om een uitspraak te doen over de effectiviteit.

In welke situaties zijn RCTs mogelijk **niet haalbaar of wenselijk** voor het aantonen van de effectiviteit van digitale zorg?

- 4) De deelnemers van de focusgroepen hebben een aantal uitdagingen genoemd m.b.t. de wenselijkheid en/of haalbaarheid van RCTs voor het aantonen van effectiviteit van digitale zorg.
- De digitale zorg is vaak al beschikbaar in de dagelijkse praktijk, waardoor patiënten zich niet willen laten randomiseren.

- Digitale zorg wordt continu doorontwikkeld met bijbehorende software updates, waardoor de resultaten van een RCT niet representatief meer zijn voor de actuele versie van de interventie.
- De effectiviteit van digitale zorg wordt sterk beïnvloed door de specifieke context waarin het wordt gebruikt. Deze context komt niet altijd overeen met de context waarin een RCT wordt uitgevoerd. Daarom is een RCT niet representatief.
- Digitale zorg wordt regelmatig ontwikkeld door startups en mkb-ondernemingen. Zij hebben, in tegenstelling tot farmaceuten, niet altijd de (financiële) middelen beschikbaar om een RCT te bekostigen.
- Nederland staat voor grote maatschappelijke opgaven qua gezondheidszorg (denk aan beschikbaarheid personeel, kosten) en het uitvoeren van RCTs voor elke digitale zorginnovatie kost (te) veel tijd.

In welke mate herkent u bovenstaande bevindingen? Wilt u nog iets toevoegen?

- 5) Digitale zorg is een verzamelnaam voor een breed en heterogeen (zorg)aanbod, zoals communicatietechnologie, telemonitoring en digitale behandelingen (*'digital therapeutics'*).

Op welke manier dient rekening te worden gehouden met, of onderscheid gemaakt te worden tussen, verschillende vormen van digitale zorg als het aankomt op de wenselijkheid en haalbaarheid van RCTs voor het aantonen van effectiviteit?

- 6) Een alternatief voor bewijslast uit RCTs is evidence verkregen uit observationele studies. Dit is veelal makkelijker te verzamelen, minder tijdrovend, minder duur en daarmee vaak een aantrekkelijk alternatief. Tegelijkertijd is de kwaliteit van het bewijs bij aanvang lager, waardoor causale effectiviteit aantonen lastiger is. Soms zijn er factoren waardoor de kwaliteit van bewijs omhoog gaat (dit wordt 'upgraden' genoemd).

Hoe kijkt u aan tegen het gebruik van data uit observationele studies bij de beoordeling van de effectiviteit van digitale zorg? Waar liggen kansen of juist knelpunten?

- 7) Soms wordt gepleit voor het sneller (voorwaardelijk) toelaten van digitale zorg en minder te sturen op bewijs uit klinische studies, maar aan te sturen op het verzamelen van 'real world evidence' (RWE). Digitale oplossingen kunnen vaak makkelijk grote hoeveelheden RWE verzamelen.

Hoe kijkt u aan tegen het gebruik van RWE bij de beoordeling van de effectiviteit van digitale zorg?

- 8) Tijdens de focusgroepen werden onderstaande uitdagingen genoemd t.a.v. het gebruik van RWE bij de beoordeling van effectiviteit van digitale zorg.
- De benodigde epidemiologische en statistische methoden om de causale effectiviteit uit (grote) observationele datasets bloot te leggen zijn zeer complex
 - De benodigde data infrastructuur om registratie, toegankelijkheid, en privacy te waarborgen ontbreekt
 - Vaak is sprake van suboptimale en/of ontbrekende data
 - De digitale vaardigheden vereist voor adequate dataverzameling ontbreken

In welke mate herkent u bovenstaande bevindingen? Wilt u nog iets toevoegen?

- 9) Hoe kijkt u naar de potentiële meerwaarde van continue dataverzameling d.m.v. RWE om zo post-marketing monitoring of cyclische beoordeling van digitale zorg mogelijk kan maken?
[Tekstvak]

In het bestaande SWP-beoordelingsproces wordt voorafgaand aan het verzamelen van literatuur op systematische wijze een inschatting gemaakt van welke studieopzet (bijvoorbeeld een RCT of prospectieve observationele studie) wenselijk en haalbaar is gegeven het doel en de inzet van digitale zorg interventie. Dit wordt *passend onderzoek* genoemd, dat in kaart wordt gebracht aan de hand van de *passend onderzoek vragenlijst*.

Deze *passend onderzoek vragenlijst* bevat de onderdelen die in onderstaande figuur beknopt staan samengevat. Meer informatie en toelichting bij elk van de vragen is terug te vinden in deze publicatie.²⁸

Pakketsbeheervragen

- Is er sprake van een **aanpassing** van een (eerder beoordeelde) interventie?
- Is er sprake van een **indicatieverruiming** *en* is er een plausibele **pathofysiologische** verklaring?

Validiteitsvragen

- Is er sprake van **natuurlijk beloop**?
- Is er mogelijk sprake van **confounding by Indication**?
- Is er een **alles-of-niets direct effect** *en* is er een plausibele **pathofysiologische** verklaring?
- Zijn de uitkomstmaten **subjectief**?
- Is er mogelijk sprake van een sterke **patiëntvoorkeur** voor de interventie?
- Is er mogelijk sprake van een sterke **behandelaarsvoorkeur** voor de interventie?

Haalbaarheidsvragen

- Zijn uitkomsten pas op **lange termijn** inzichtelijk?
- Is er mogelijk sprake van **patiënt-equipose**?
- Is er mogelijk sprake van **klinische equipose**?
- Is er sprake van een **zeer klein indicatiegebied**?
- Is deze interventie de **enige behandelmogelijkheid** voor de doelpopulatie en mag deze niet onthouden worden? (bijv. zeer slechte prognose of kwaliteit van leven)
- Is er **hoge urgentie** bij het indicatiegebied?
- Heeft de interventie een **hoog multifactorieel karakter**?
- **Aanvullende redenen** die onderzoekskenmerken moeilijk of onmogelijk maken

²⁸ Heymans JM. Passend bewijs bij het bepalen van effectiviteit van interventies. Ned. Tijdschr. Geneeskd.

- 10) Is de passend onderzoek vragenlijst wat u betreft bruikbaar en geschikt voor digitale zorg? Licht uw antwoord toe.
- Ja *[ruimte voor toelichting en toevoegingen]*
 - Nee *[ruimte voor toelichting]*
- 11) Ontbreken er vragen op de passend onderzoek vragenlijst of zijn er bepaalde vragen overbodig bij beoordeling van digitale zorg?
- Nee
 - Ja, te weten ... *[ruimte voor toelichting]*

Overige overwegingen ('from evidence to decision')

Naast het oordeel over effectiviteit kunnen ook andere argumenten een rol bij het beantwoorden van de vraag of de interventie thuishoort in het verzekerde pakket.

Zo kunnen criteria als **kosteneffectiviteit**, **noodzakelijkheid** en **uitvoerbaarheid** worden meegewogen bij de beslissing of een interventie moet worden opgenomen in het basispakket.

Tijdens focusgroepen werden de volgende aanvullende argumenten genoemd:

- toegankelijkheid
- belasting zorgprofessional
- houdbaarheid zorgsysteem

- 12) Welke van bovenstaande overige overwegingen vindt u belangrijk om mee te nemen in de beoordeling van digitale zorg? Heeft u nog aanvullende overwegingen?

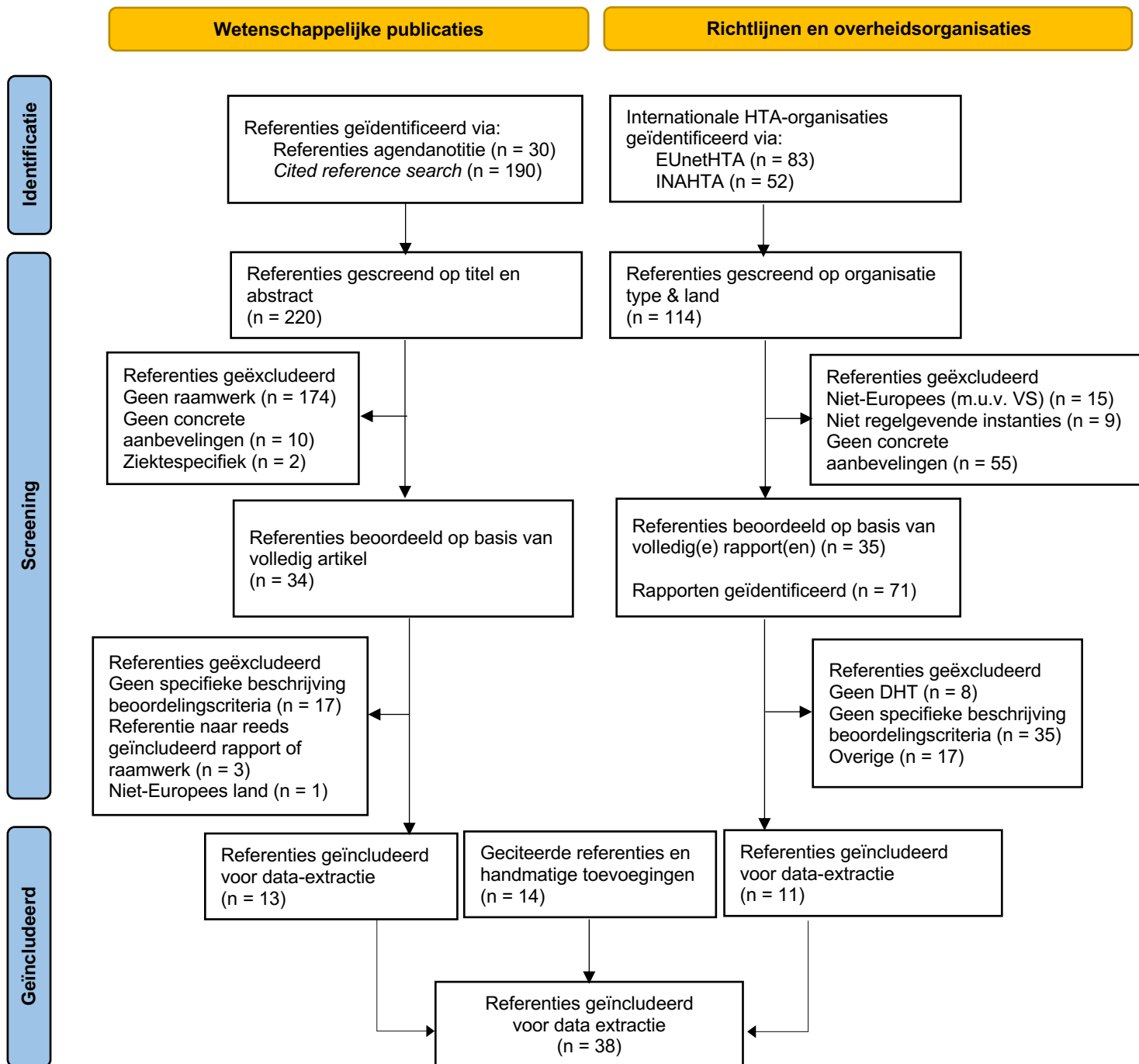
[Tekstvak]

- 13) Hoe zwaarwegend moeten dergelijke argumenten wat u betreft zijn? Moet kosteneffectiviteit of personele inzetbaarheid bijvoorbeeld doorslaggevend zijn, ook als blijkt dat digitale zorg minder effectief is dan de bestaande standaardzorg?

[Likert schaal voor elk van bovenstaande 6 criteria.]

- 1 Niet van belang
- 2 Minder belangrijk
- 3 Belangrijk
- 4 Zeer belangrijk
- 5 Essentieel

Bijlage 3. Flowchart verkennend literatuuronderzoek



Bijlage 4. Beschrijving beoordelingsprocedure per instantie en land

4.1 Federal Institute for Drugs and Medical Devices (BfArM), Germany

Scope: Low-risk medical device (class I and IIa CE)

Abbreviations

DiGA Digitale Gesundheitsanwendungen (=Digital Medical Technologies)
DMTs Digital Medical Technologies
MD Medical device

Identification

No information.

Selection / eligibility criteria

In order to be eligible as a DMT, an application should have the following characteristics:

1. Low-risk medical device
2. The DMT is used as a support to
 - a. the identification, monitoring, treatment or alleviation of disease
 - b. the identification, treatment or alleviation or compensation of injuries or disabilities
3. Its main function should be based on digital technologies
In principle, a DiGA can be an application on a smartphone as well as a desktop or browser-based application. A DiGA can also comprise devices, sensors, or other hardware in addition to software, such as wearables, as long as the main function is a predominantly digital one, the hardware is necessary to achieve the purpose of the DiGA and the hardware is not a privately financed item of everyday life such as a gym mat or smartphone for doing the exercises guided by the DiGA
4. Its main function should not merely consist in the collection of data
It is not sufficient for the DMT to present the collected data to the user, it has to elaborate the data to extract new original information (e.g. treatment suggestion).
5. It should be used autonomously by the patient or by the patient together with the healthcare professional

Not included:

- primary prevention of a disease, e.g. DMTs designed to promote a healthy lifestyle or to prevent the insurgence of a pathology are not considered as DiGA, as these are not considered medical devices. (secondary and tertiary prevention are included)
- Telemedicine and telemonitoring are not included as it requires significant input from a healthcare professional to achieve the medical purpose.

Evaluation

In order to be eligible for reimbursement, the DMT has to fall within the definition of DiGA and has to pass the evaluation of the BfArM (the precise conditions are described below). Furthermore, there must be either a prescription of a physician/psychotherapist or an approval of the health insurance company.

The DMT manufacturer must include proof of the following in their DMT to be included in the directory:

1. Safety and suitability for use
The safety and functional reliability should be confirmed by the CE mark.

2. Interoperability

This is defined as the ability of technical systems to cooperate on a technical-syntactical, semantic, and organizational level. The DMT must meet certain interoperability requirements:

- The DiGA allows the insured person to export therapy-relevant extracts of the data collected via the DiGA in human-readable and printable form, so that he can use them for his own purposes or pass them on to a physician.
- The DiGA allows the insured person to export the data collected from the DiGA in a machine-readable, interoperable format so that the insured person or a third party authorised by the insured person can further process these data via other digital products.
- If the DiGA obtains data from medical devices used by the insured person or sensors worn by the insured person for the measurement and transmission of vital signs (wearables), it may also address these devices via an interoperable interface.

3. Further quality requirements

- Robustness: against incidents and user errors
- Consumer protection: meet consumer law requirements
- Ease of use
- Support for healthcare providers: a clear description of the role of the healthcare providers, their interaction, the legal requirements
- Quality of medical content: conform to accepted medical standards
- Patient safety: include measures aimed at ensuring patient safety

4. Data protection requirements and information security according to the state of the art

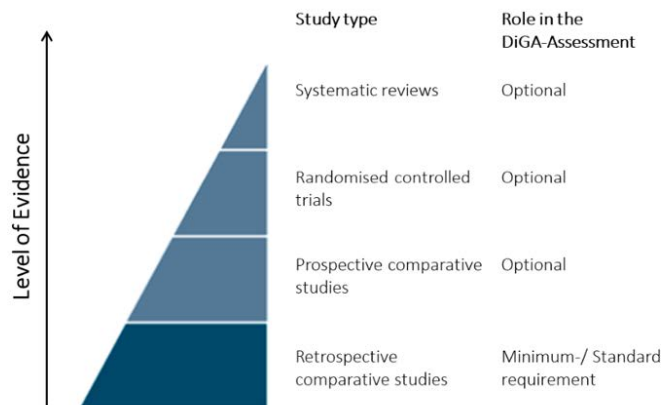
5. Evidence of “positive healthcare effects”

- either a medical benefit: morbidity, mortality, or quality-of-life:
 - the improvement of the state of health,
 - the reduction of the duration of a disease,
 - the prolongation of survival or
 - an improvement in quality of life (can be demonstrated based on disease specific as well as generic scales).(comparable to endpoint used in the benefit assessment of pharmaceuticals)
- or a patient-relevant improvement of structure and processes in health care: aimed at supporting the health behaviour of patients or integrating the processes between patients and healthcare providers. This includes the areas of:
 - coordination of treatment procedures,
 - alignment of treatment with guidelines and recognised standards,
 - adherence (on the basis of e.g. Morisky Medication Adherence Scale),
 - facilitating access to care (on the basis of e.g. average waiting time until start of therapy, or therapy discontinuation rates),
 - patient safety,
 - health literacy (on the basis of e.g. European Health Literacy Survey Questionnaire (HLS-EU-Q),
 - patient autonomy,
 - coping with illness-related difficulties in everyday life, or
 - reduction of therapy-related efforts and strains for patients and their relatives.

The demonstration of a positive healthcare effect on one of the outcomes previously mentioned is sufficient. If the effect on multiple outcomes is demonstrated, then a statistical analysis on multiplicity

is required, which may complicate the analysis. Therefore, in practice BfArM often recommends focusing on one outcome only.

A quantitative comparative study is required which shows that using DiGA is better than not using it (can be a prospective comparative study, RCT or also a retrospective intra-individual comparison, demonstrating the benefit of using the DMT compared to not using it). A single such study is sufficient. The required study design does not differ depending on the risk category or functionality of the DMT. Results must be statistically significant and clinically relevant.



Source: [KCE Report 362 \(2023\)](#)

If the DMT includes a diagnostic instrument, the manufacturer must also submit a study on the sensitivity and specificity of the DMT in the targeted patient group.

The study must be conducted in Germany or, if the study has not been conducted in Germany, the manufacturer must prove that its results are applicable in the context of the German healthcare system.

Complete study results have to be published. Initially, publication on the manufacturer's website is sufficient; later, the study must be included in a public study registry. An additional publication in a peer reviewed journal is not necessary but can be advantageous in terms of professional reputation and acceptance.

If the manufacturer cannot prove the positive healthcare effects yet, the DMT may be preliminarily included in the registry (as "registration for evaluation") for up to 12 months, (which can be extended to 24 months if well justified), during which temporary reimbursement is provided. Such "registration for evaluation" requires the manufacturers to:

- describe the expected positive healthcare effects
- submit a scientific evaluation from an independent third-party institution, that discusses the positive effects and which must include at least a statistical data analysis of the use of the DMT. The statistical data analysis can be based on an intra-individual comparison, non-randomised comparison (historical control group), randomised pilot study or intermediate results of a randomised study. It must include a systematic literature review, a description of the design and methods and a results and discussion section.

Changes to the DiGA which have a significant influence on the evaluation decision of the BfArM, or which may lead to changes in the information in the directory must be reported as such to the BfArM. The manufacturer can assess whether a change to the DiGA is a "significant change" with a checklist on the [BfArM website](#).

Decision-making (i.e. reimbursement or implementation)

Whereas the BfARM is responsible for evaluation and assessment, it is the GKV-SV (Spitzenverband der Gesetzlichen Krankenkassen – Head Association of the Statutory Health Insurances) which is responsible for the price negotiations for the pricing of the DMT. For the first 12 months the price is freely set by the manufacturer. Only after this period the price is negotiated with the GKV-SV. If no agreement is reached in the negotiations, an arbitration board sets the reimbursement price.

A lot of DMTs fall outside the scope of the fast-track procedure, e.g. products for monitoring of a pacemaker or an infusion pump, which are classified as Class III products. These DMTs have to follow the usual pathway for reimbursement by the statutory health insurance.

The process for DMTs that are not DiGAs follows the methods defined in the [Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen \(IQWiG\) General Methods document](#).

Besides the DiGA pathway, Germany offers other potential pathways for (higher risk) DMTs to be reimbursed:

- As outpatient care, DMTs can be reimbursed under a physician-procedure code
- For inpatient care, DMTs can be classified under the Diagnostic Related Groups (DRG) system or as new treatment; alternatively it can be regarded as infrastructure investment
- It can be financed through the innovation fund, which promotes new forms of care that exceed the current standards. Note that this fund covers all types of innovations in general, not just digital health innovations.
- It can be financed on the basis of selective payer contracts: Disease Management Programs (DMPs) or integrated care contracts. Most telemedicine projects are currently being covered and tested in the latter ones (e.g. telemonitoring for congestive heart failure, video Parkinson therapy).
- In case of disease prevention, it can be reimbursed from the statutory health insurance disease-prevention budget

Additional information & support

In 2020 the Federal Institute for Drugs and Medical Devices (BfArM, Bundesinstitut für Arzneimittel und Medizinprodukte) launched a fast-track procedure for the reimbursement of DMTs to facilitate their way into regular care. The reimbursed DMTs are registered in a specific Directory (Digitale Gesundheitsanwendungen, DiGAs), which can be directly consulted [online](#).

With this fast-track procedure, manufacturers can obtain either temporary or permanent reimbursement. The submitted files are evaluated in a time frame of 3 months. Statutory health insurance reimburses the full price of the DMT; there is no co-payment by the patient.

Sources

- [KCE Reports 362 Evaluation of Digital Medical Technologies](#)
- https://www.bfarm.de/EN/Medical-devices/Tasks/DiGA-and-DiPA/Digital-Health-Applications/_node.html

4.2 Food & Drug Administration (FDA), United States

Scope Medical devices (including software)

Abbreviations

CDRH Center for Devices and Radiological Health
FDA Food & Drug Administration
DHCoE Digital Health Center of Excellence

Identification

There are four steps to bring a device to the market:

- [Step One: Classify Your Device and Understand Applicable Regulatory Controls](#)
- [Step Two: Select and Prepare the Correct Premarket Submission](#)
- [Step Three: Send your Premarket Submission to the FDA and Interact with FDA Staff During Review](#)
- [Step Four: Comply with Applicable Regulatory Controls Including the Establishment Registration and Device Listing](#)

These steps are explained in the sections below.

Selection / eligibility criteria

The following steps may be helpful for determining whether a product is regulated by the FDA as a medical device.

[Step 1: Determine if your product meets the definition of a medical device](#)

[Digital Health Center of Excellence \(DHCoE\)](#) considers DHTs as part of medical devices, where ‘device’ is defined per Section 201(h) of the Food, Drug & Cosmetic Act. as: *an instrument, apparatus, implement, machine, contrivance, implant, in vitro reagent, or other similar or related article, including a component part, or accessory, which is, among other criteria, intended for use in the diagnosis of disease or other conditions, or in the cure, mitigation, treatment, or prevention of disease in man.*

[Step 2: Determine if an appropriate product classification exists for your product](#)

Medical devices are [categorized](#) into one of three classes (I, II, or III), based on the degree of risk they present. As device class increases from class I to class II to class III, the regulatory controls also increase, with class I devices subject to the least regulatory control, and class III devices subject to the most stringent regulatory control. The classes of devices, regulatory controls and submission types are summarized in the table.

Device classification depends on the *intended use* of the device and also upon *indications for use*. For example, a scalpel's intended use is to cut tissue. A subset of intended use arises when a more specialized indication is added in the device's labeling such as, "for making incisions in the cornea".

Indications for use can be found in the device's labeling, but may also be conveyed orally during sale of the product.

Class	Risk	Potential Harm	Regulatory Controls	Submission Type or Exemption
I	Lowest	Present minimal potential for harm	General	510(k) 510(k) Exempt
II	Moderate	Higher risk than class I devices	General and Special (if available)	510(k) 510(k) Exempt
III	Highest	Sustain or support life, are implanted, or present potential unreasonable risk of illness or injury	General and PMA	PMA

There are several options to explore whether a product classification already exists:

1. Search the [FDA Product Classification Database](#) to determine if there is an existing product classification that applies to your product
2. Search for Similar Devices with market authorization
 - a. [Premarket Notification 510\(k\)](#)
 - b. [Premarket Approval \(PMA\)](#)
 - c. [De Novo](#)
 - d. [Humanitarian Device Exemption \(HDE\)](#)
3. Search for Similar Devices by [Device Listing](#)
 - a. You may search for a legally marketed device's product classification by reviewing its device listing information. Device listing information can be found by searching the FDA's

Specifically for software, [The Digital Health Policy Navigator](#) provides considerations for whether a software function is potentially subject to or the focus of FDA's regulatory oversight as a device. The Digital Health Policy Navigator does not address:

- Considerations on how to design, develop, and test software functions; or
- Considerations relating to the regulation of software used with other medical products (e.g., drugs, biologics), or software intended for other purposes (e.g., electronic consent forms).

If no classification exist you need to establish the classification for your specific product. The following resources may assist you in determining the classification of your product:

- [FDA Product Classification Database](#)
- [Device Classification Panels](#)
- [Classify Your Medical Device](#)
- [Medical Device Accessories](#)

If your product is a combination product - a medical device plus another FDA-regulated product (e.g. drug, biologics, etc.), you should contact FDA's Office of Combination Product (OCP)

Exemptions

Some class I and class II devices are [exempt](#) from the 510(k) notification requirement, if they do not exceed the limitations of exemption stated in 21 CFR xxx.9, where xxx refers to 21CFR 862-892. For

example, an elastic bandage classified under 21CFR 880.5075 is exempt from premarket notification, provided that it does not exceed the exemption limitations stated in 21CFR 880.9.

Approximately 572 (or 74%) of the Class I devices are exempt from the premarket notification process. These exemptions are listed in the classification regulations of 21 CFR and also has been collected together in the [Medical Device Exemptions](#) document.

Evaluation

A device facility must register its establishment and list its devices with the FDA. Owners or operators of establishments that are involved in the production and distribution of medical devices intended for use in the U.S. are required to register annually with the FDA. This process is known as establishment registration. Information about how to register and list may be found at [Device Registration and Listing](#).

If a device requires a premarket submission prior to marketing (i.e., the medical device is not exempt), the device facilities must wait until it receives FDA clearance or approval before registering and listing.

Registration of a device establishment, assignment of a registration number, or listing of a medical device does not in any way denote clearance or approval of the establishment or its products by the FDA.

The [requirements](#) for a specific DHT depend on the class it is in.

[Classes](#)

Class 1: General Controls

Class 1 devices pose the least amount of risk to consumers. These low-risk devices, such as oxygen masks or surgical tools, are subject to “general controls.” General controls ensure the safety and effectiveness of devices once they’re manufactured. General controls consider the following factors:

- [Good Manufacturing Practices](#)
- Standards and Reporting Adverse Events to FDA
- [Register their establishments and list](#) the medical devices they market with FDA
- General recordkeeping requirements
- Are in accordance with [labeling regulations](#) and cannot be [adulterated or misbranded](#).

A full list of general controls can be found [here](#).

Class 2: General Controls With Special Controls

Class 2 devices pose more risk to consumers than do Class 1 devices. Therefore, Class 2 devices are subject to special controls in addition to general controls. Special controls include:

- Labeling requirements (information that must be included on a product label)
- Device specific mandatory performance standards
- Device specific testing requirements

Class 3: General Controls and Premarket Approval

Usually, Class 3 devices support or sustain life, are implanted in the body, or have the potential for unreasonable risk of illness or injury. Examples include pacemakers, breast implants, and HIV diagnostic tests. As a result, Class 3 devices require premarket approval. To receive this, a manufacturer must prove that a device is safe and effective. Class 3 devices are also subject to general controls.

Submission type

If medical device developers have enough information on a device's safety and effectiveness, they can file an application to market the device to the public. The type of application they file depends on the device's class.

The following submission types are distinguished:

1. Premarket Notification (510(k))

Premarket Notification, also known as a 510(k), indicates that the Class 2 medical device is similar to others on the market. To support the claim, the developer compares the new device to one or more similar, legally marketed devices.

A device is considered *substantially equivalent* if it has the same intended use, technological characteristics, and performance testing (only Class II) as a legally marketed device. A legally marketed device was:

- legally marketed prior to May 28, 1976 ("preamendments device"), for which a PMA is not required, or
- reclassified from Class III to Class II or Class I, or
- found substantially equivalent through the 510(k) process.

Applicants must compare their device to one or more similar legally marketed devices and make and support their substantially equivalent claims. If the device is substantially equivalent to an approved medical device, it is placed in the same class. If it is not substantially equivalent, it becomes non-SE and is placed into Class III.

Examples of 510(k)s include x-ray machines, dialysis machines, fetal monitors, lithotripsy machines, and muscle stimulators.

2. Premarket Approval (PMA)

Premarket Approval applications must be submitted for Class 3 devices and must include data from all nonclinical studies and clinical studies. During the approval process, FDA will inspect the manufacturing laboratories and facilities where the device will be made to check for good manufacturing practices.

The sponsor must provide valid scientific evidence demonstrating reasonable assurances of safety and effectiveness for the device's intended use.

Class III devices require a PMA. A PMA is the most stringent type of premarket submission. Before the FDA approves a PMA, the sponsor must provide valid scientific evidence demonstrating reasonable assurances of safety and effectiveness for the device's intended use.

PMA refers to the scientific and regulatory review necessary to evaluate:

- the safety and effectiveness of Class III devices or
- devices that were found not substantially equivalent to a Class I or II predicate through the 510(k) process.

PMA is the most involved process. To reasonably determine that a device is safe and effective. PMA requires:

- scientific evidence that the possible benefits to health from the intended use of a device outweigh the possible risks
- that the device will significantly help a large portion of the target population.

Independence is an important concept for PMAs, meaning that each PMA should establish the safety and effectiveness of the device under review, and that data about one device cannot be used to support another.

Examples of PMAs include digital mammography, minimally invasive and non-invasive glucose testing devices, implanted defibrillators, and implantable middle ear devices.

3. De Novo Classification Request

General controls alone, or general and special controls, provide reasonable assurance of safety and effectiveness for the intended use, but for which there is no legally marketed predicate device

The De Novo process provides a pathway to classify novel medical devices for which general controls alone, or general and special controls, provide reasonable assurance of safety and effectiveness for the intended use, but for which there is no legally marketed predicate device.

4. Humanitarian Device Exemption (HDE)

HDE intended to benefit patients with rare diseases or conditions. In order for a device to be eligible for an HDE, a sponsor must first obtain designation as a Humanitarian Use Device (HUD), which is granted through an application to FDA's Office of Orphan Products Development (OOPD).

Humanitarian Use Devices benefit patients by treating or diagnosing a disease or condition that affects fewer than 4,000 people. Before they can market a Humanitarian Use Device, developers must submit a human device exemption and must demonstrate that there are no similar, legally approved devices on the market and that there is no other way to bring a Humanitarian Use Device to market.

Evidence guidance

Design Controls: Class II and class III devices must be designed in accordance with Design Controls under the Quality System Regulation ([21 CFR 820.30](#)). Most class I devices are not subject to Design Controls. For additional information on Design Controls, please see: [Design Control Guidance for Medical Device Manufacturers](#).

Nonclinical Testing: The types of information and testing necessary to market your device are determined by many factors, including the device classification, principles of operation, technological characteristics, and labeling. Nonclinical testing performed in support of a premarket submission for a medical device should comply with the Good Laboratory Practices (GLPs) in [21 CFR 58](#). For additional information please see [Recommended Content and Format of Complete Test Reports for Non-Clinical Bench Performance Testing in Premarket Submissions](#).

Consensus Standards: The FDA encourages the use of FDA-recognized consensus [standards](#) in premarket submissions. An integral part of a least-burdensome approach to device review is a reliance upon international consensus standards. Conformance is voluntary, unless a standard is incorporated by reference into regulation (see the [Standards Medical Devices Federal Register Documents](#)). The guidance, [The Appropriate Use of Voluntary Consensus Standards in Premarket Submissions for Medical Devices](#), explains how to use standards in device submissions and outlines the documentation needed to support their use, including declarations of conformity and test reports.

The FDA recommends that manufacturers consider participation in the voluntary [Accreditation Scheme for Conformity Assessment](#) (ASCA) Program. This voluntary pilot program is designed to decrease the burden of individual premarket submissions by enhancing consistency and predictability in the FDA's approach to assessing conformance with ASCA-eligible standards in medical device premarket reviews. The FDA has issued three final guidance documents related to the voluntary ASCA Program which can be downloaded from the [ASCA web page](#).

Clinical Evidence: PMAs, HDEs and some 510(k)s and De Novo Classification Requests require clinical evidence. Prior to initiating a clinical study, the study sponsor may need to obtain approval of an Investigational Device Exemption (IDE) by the FDA. The study will also need to be approved by the appropriate Institutional Review Board (IRB). Clinical studies must comply with all applicable IDE regulations and Good Clinical Practices (GCPs). For additional information on the IDE and GCP regulations, see [Investigational Device Exemption \(IDE\)](#).

Labeling: The labeling for a device must be written according to labeling regulations and included in your premarket submission. For information on Device Labeling and UDI requirements please refer to the following references:

- [21 CFR 801 Labeling](#)
- [Device Labeling](#) and [Unique Device Identification System \(UDI\)](#)
- [Global UDI Database \(GUDID\)](#)
- [CDRH Learn \(UDI Modules\)](#)

Decision-making (i.e. reimbursement or implementation)

Once you have prepared the appropriate premarket submission for your device, you need to send your submission to the FDA and interact with FDA staff during its review. Once your submission has been received by the FDA, you should be aware of the following:

- **Administrative Review:** After a premarket submission is received, the FDA conducts an administrative review to assess whether the submission is sufficiently complete to be accepted for substantive review. Please see the [Refuse to Accept Policy for 510\(k\)s](#), [Acceptance and Filing Reviews for Premarket Approval Applications \(PMAs\)](#) or [Acceptance Review for De Novo Classification Requests](#) and [Humanitarian Device Exemption \(HDE\) Program: Guidance for Industry and Food and Drug Administration Staff](#).

- **Interactive Review:** While a submission is under review, FDA staff communicates with applicants to increase the efficiency of the review process. Please see: [Types of Communication During the Review of Medical Device Submissions](#).
- [Send and Track Medical Device Premarket Submissions Online: CDRH Portal](#) - When you send a CDRH 510(k) submission (traditional, special, and abbreviated 510(k)s) for review, your official correspondent can monitor the FDA's progress online in a simple, concise format.

Additional information & support

[Division of Industry and Consumer Education](#) (DICE) or DeviceDetermination@fda.hhs.gov if you are unable to make a device determination for your product.

The FDA has developed several different types of resources to help you prepare a premarket submission. These include the following:

- [Send and Track Medical Device Premarket Submissions Online](#): Register for a CDRH Portal account to send CDRH eCopy or eSTAR premarket submissions online
- [CDRH Learn](#): A video-based series of teaching modules, seminars, and recorded webinars that cover a variety of policy and guidance efforts.
- [CDRH's Q-Submission Program](#): Prospective applicants of future premarket submissions may request feedback from the FDA through a Pre-Submission, which is part of the Q-Submission Program. We encourage you to read our online information and other available resources prior to sending any request for feedback.
- [CDRH Breakthrough Devices Program](#): This voluntary program helps patients have more timely access to certain medical devices that treat or diagnose a life-threatening or irreversibly debilitating condition by speeding up their development, assessment, and review. Devices subject to Premarket Approval applications (PMAs), premarket notifications [510(k)] or De Novo classification requests may be eligible. Sponsors can send a Breakthrough Designation request for their device at any time prior to sending the device's marketing submission.

Source

[Digital Health Center of Excellence | FDA](#)

4.3 Austrian National Public Health Institute (Gesundheit Österreich GmbH, GOeG), Austria

Scope: CE marked Digital Medical Technologies (DMTs; not limited) classified as DMT and available in German. Focus on therapeutic DMTs for chronic illnesses/patients. Diagnostics excluded due to potential risks linked to patient self-diagnosis

Abbreviations

AIHTA Austrian Institute for Health Technology Assessment
DMTs Digital Medical Technologies
MD Medical device

Note

A short description of the Finnish Digi-HTA framework is provided at the end of the Austrian process description for DHTs, as this was used as the basis for the Austrian system.

Identification

Most applications procedures for the reimbursement of DMTs are likely to come from manufacturers directly, but DMTs developed in collaboration with the healthcare insurances (sickness funds) would also be considered.

Selection / eligibility criteria

A prioritisation process was considered necessary; relevant criteria (established via expert consultations):

- DMTs (CE marked) categorised as such and be available in German.
- DMTs with a therapeutic purpose highlighted as a priority. Diagnostics excluded due to potential risks linked to patient self-diagnosis. Similarly primary prevention excluded as it falls out of the statutory health insurance benefits under the Austrian General Social Insurance Act (ASVG).
- Specific focus on chronic illnesses for two reasons:
 - these were thought to be the patients that could use these DMTs over a longer time frame, benefiting the most from their integration in clinical care pathways, which could in turn result in clinical benefits or organisational benefits (e.g. a reduction in the consumption of healthcare resources).
 - for acute or emergency situations there are currently very few apps, so prioritising use in chronic patients seemed logical as a starting point (could, should evolve and be adapted over time)
- All types of DMTs could be considered in this framework, both DMTs for self-use by the patient or those to be used by the patient in consultation or collaboration with the healthcare professional. DMTs to be used exclusively by healthcare professionals are excluded, since these are covered by a different budget in Austria (e.g. a specific software for exclusive use of healthcare providers).

Evaluation

Requirements for the reimbursement of DMTs were classified into three groups:

- *Technology specific* requirements, such as data privacy or technical standards for interoperability or algorithms.
- *Regulatory* requirements, including CE marking and risk class, information often captured in the web page of most manufacturers. Also under this category of requirements, the

classification of each DMT in tiers was made. This classification made use of the tiers used and described in the NICE ESF for DMT.

- *Evidence* requirements were set mirroring those of the NICE ESF (Evidence Standards Framework) for DMTs. In this step, the minimum requirements as well as gold standards would be considered for any new DMT being evaluated in order to ensure that enough evidence exists and supports the reimbursement and use of the DMT. For example, for higher risk DMTs, RCTs would be requested, while for lower risk DMTs, this may be a nice to have, but not always necessary. Comparators or study types would be looked at and if considered inadequate, the DMT would not be deemed suitable for evaluation. This would require a case-by-case consideration

No economic requirements, no ethical requirements; the focus of the framework was on safety and effectiveness.

If all the necessary requirements previously listed are fulfilled, then an evaluation would be performed, and other aspects would then be expected to be covered.

In order to appropriately do so, two further frameworks were recommended to be used during this evaluation:

- the Digi-HTA framework used in Finland*
- HTA module for mobile health applications (MMA) developed by [Moshi et al.](#) (theoretical framework applied to Australian context)

These two frameworks ensure all key points are thought about, and covered during reimbursement discussions (and nothing crucial is missed). They do not respond to a standardised approach, by which if a specific question is not answered, the DMT would be given a negative recommendation. Instead, they serve more as guidance.

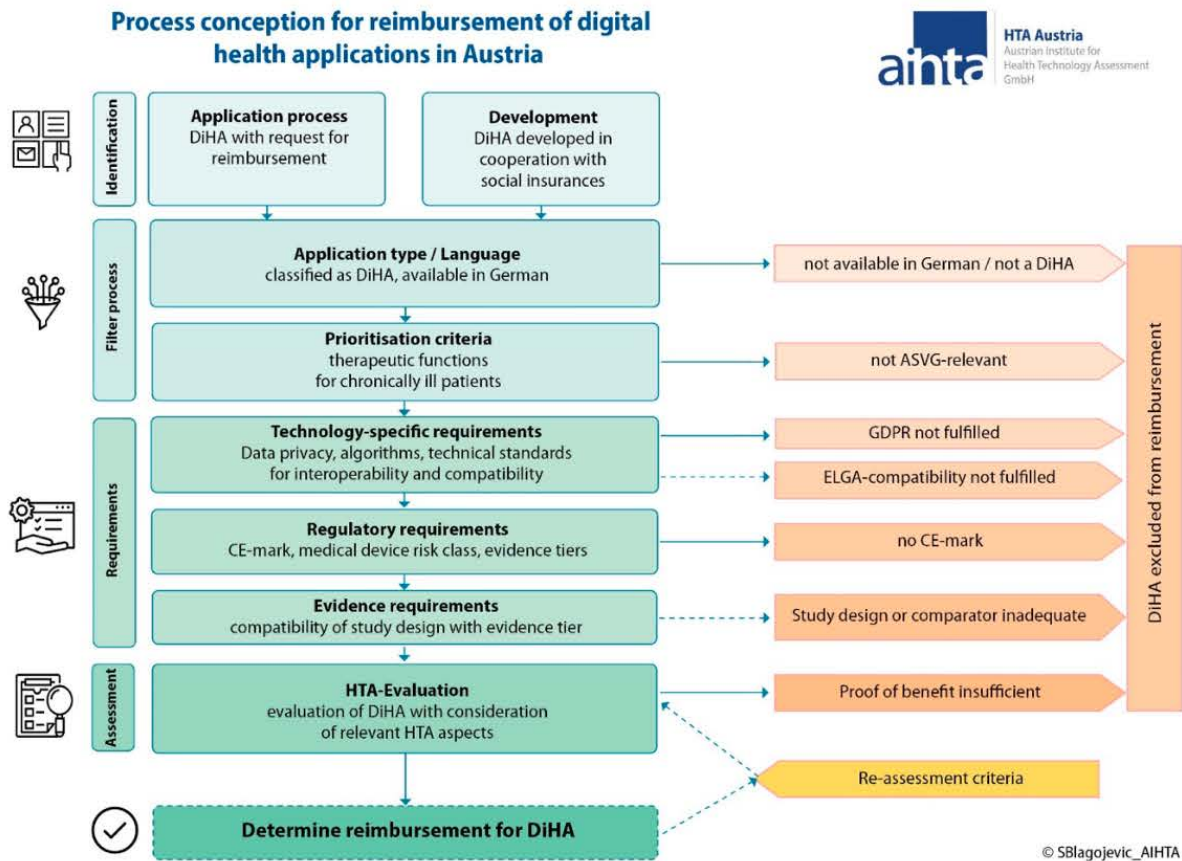
Decision-making (i.e. reimbursement or implementation)

The pathway for price setting for DMTs has not yet been decided. Free pricing even for a limited period of time, was thought not to be appropriate, since prices may in such case be high and result in a high budget impact.

Additional information & support

A framework for the evaluation of DMTs in Austria is currently under development.

The KCE report, which was used as a source for this summary, based its description on all the research published by the AIHTA (Austrian Institute for Health Technology Assessment) in this field up to date and discussions with one of the authors of these articles, who was involved in the development of this framework from its origin. The procedure described has not yet been implemented, and remains to this date, a theoretical framework.



Source: [KCE Report 362 \(2023\)](#)

*Digi-HTA framework, Finland

As a traditional HTA does not cover all factors relevant to DMTs, an adapted framework for DMTs was developed in Finland, named Digi-HTA. The nature of this HTA is at present only voluntary and its recommendations remain purely informative.

Digi-HTA assessments are done by the Finnish Coordination Centre for Health Technology Assessment (FinCCHTA), mainly based on information provided by the company about its own product, supplemented by a literature HTA review done by HTA experts.

With the Digi-HTA recommendations, municipalities (n=300) are informed about the suitability of the use of a digital healthcare service. (Healthcare in Finland consists of a highly decentralised publicly funded healthcare system. The municipalities decide freely on the care provided and purchase through procurements.)

Scope of the framework: digital technologies, comprising mHealth, artificial intelligence and robotics.

Three major developmental waves in the digitalisation in healthcare are distinguished:

- First wave: basic local healthcare information infrastructure, like electronic medical records, picture archiving, communication systems, laboratory systems, other auxiliaries to medical records, regional communication, and repositories as well as referral systems
- Second wave: characterised by further utilisation of the collected and stored data, as well as citizen involvement
- Third wave: novel solutions involving AI and machine learning

The evaluation process not only requires HTA experts but also experts in data security and protection. Traditionally, the key domains in HTA recommendations have been safety, effectiveness, and cost. The Digi-HTA adds 4 extra domains:

- usability and accessibility (for both end-users and healthcare professionals)
- data security and protection
- interoperability
- technical stability

For each domain, the Digi-HTA framework lists several questions.

Two of these domains, (i.e. interoperability and technical stability), are not considered in the scoring system, as strict requirements were already in place at the time the Digi-HTA system was developed. For the remaining five domains, a scoring system feeds into traffic lights and the scores for each domain are consecutively combined in an overall evaluation.

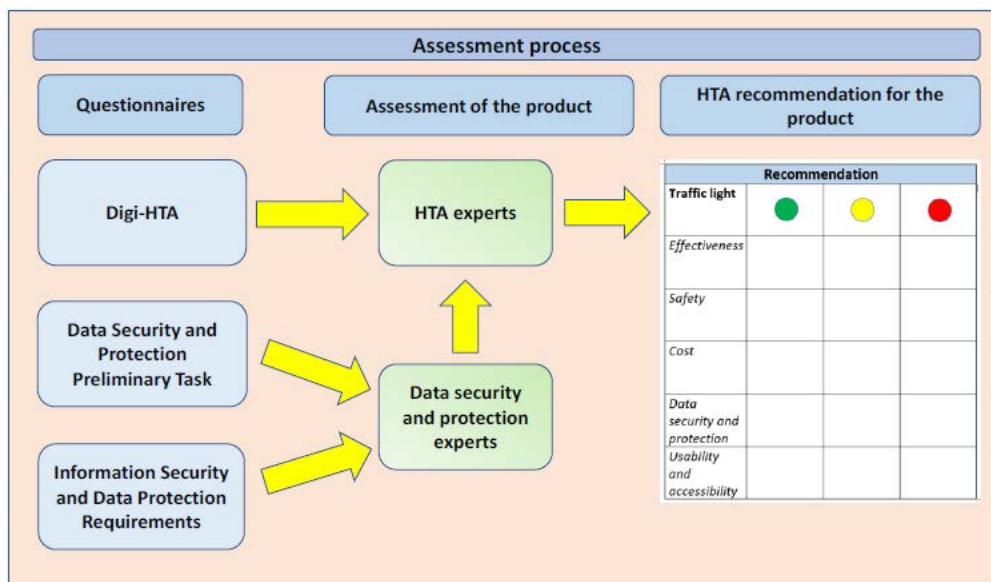


Table 11 – Scoring in the Digi-HTA assessment

Score	Effectiveness	Safety	Cost	Data security and protection	Usability and accessibility
2	Adequate	Adequate	Reasonable	Adequate	Adequate
1	Promising, but more evidence is needed	Apparently adequate, but insufficiently known	High	Minor deficiencies	Minor deficiencies
0	Bad or unknown	Bad or unknown	Prohibitively expensive	Flaws	Flaws

Table 12 – Recommendation scale

Total score	Key
10	The use of the product is recommended. The product receives a sufficient score on all domains.
9	On one of the domains, the product does not receive a sufficient score.
7-8	A number of factors (2 or 3 domains) need to be taken into account.
5-6	A large number of factors (4 or 5 domains) need to be taken into account.
0-4	Some key factors are seriously flawed and must be considered before deploying the product.

Source: ⁵⁹

Source: [KCE Report 362 \(2023\)](#)

For effectiveness high quality studies are requested, but these do not need to necessarily be a RCT (case-by-case basis, depending on the DMT being studied – e.g. the requirements would not be the same for a class I or a class III MD). If there are similar products already in the market or for which RCTs have been carried out, these are also considered (Digi-similar RCTs), since they are believed to be informative. No cost-effectiveness is included in the framework but costing questions are covered.

If an evaluation results in a score of ≤ 6 , the developer/manufacturer has the right to withdraw from the procedure. They can drop out completely (since at present it has a voluntary nature) or stop the clock for around a month to gather additional information that could help to raise the score originally received.

By default, the recommendation is valid for three years. The product will be reassessed when there is new significant new information (mostly on effectiveness or on usability and accessibility) and the company requests an evaluation.

Source

[KCE Reports 362 Evaluation of Digital Medical Technologies](#)

4.4 Federaal kenniscentrum voor de Gezondheidszorg (KCE) / mHealthBELGIUM, Belgium

Scope: Mobile health apps

Abbreviations

DMTs Digital Medical Technologies

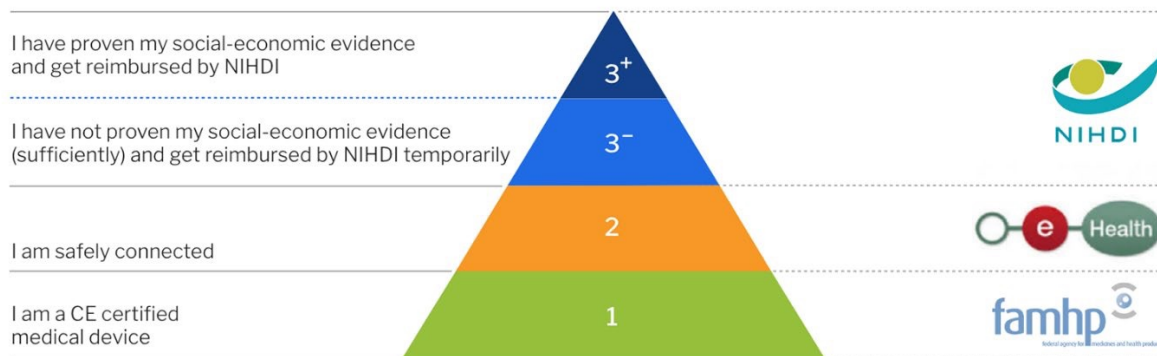
MD Medical device

Identification

Manufacturers can voluntarily register their mobile apps with assigned classification under the pyramid on the digital platform mHealthBelgium which is managed by the industry (beMedTech, the federation for manufacturers of medical devices in Belgium and Agoria, the sector federation of technological industry), with the support of the public authorities. The platform is limited to CE marked mobile health apps.

Selection / eligibility criteria

Only apps that have completed levels M1 (CE marking and GDPR) and M2 (interoperability and connectivity) of the pyramid can start the integration process in the M3 level (reimbursement).



Source: [KCE Report 362 \(2023\)](#)

The **M1 level** defines the minimal requirements for a mobile health app:

- it must be CE marked in accordance with MD legislation;
- it must be voluntary notified to the FAMHP which is the Belgian national competent authority for MD;
- the manufacturer must declare that the app complies with GDPR rules (no certification mechanism exists in Belgium).

The **M2 level** relates to interoperability and connectivity to the *eHealth platform*.

The Belgian e-Health platform is organised within the Federal Public Service for Health to promote and support well-organised mutual electronic service provision and information exchange between all healthcare actors in Belgium, with the necessary guarantees in terms of information security, protection of patient and healthcare provider privacy and respect for medical confidentiality.

In this context, [the eHealth platform develops open access and free IT services and tools](#).

According to the mHealthBelgium website, this M2 level requires the manufacturer “to have carried out a risk assessment (developed by an external organisation, service included in mHealthBelgium),

after which it will be invited to test the connectivity of its application to the core services of the e-Health platform in a test environment”.

To that end, a ‘risk assessment tool’ has been developed by IMEC (Interuniversity Micro-Electronics Center), which consists of a questionnaire that indicates which ‘automated tests’ are applicable if any. The ‘test centers’ (for automated testing of the connectivity) are however, still not implemented due to a lack of expertise within the e-Health platform to develop the appropriate architecture and because certain app manufacturers are developing their own interoperability and connectivity systems which are not always transparent. Therefore, the e-Health platform is not delivering a formal confirmation of compliance with M2.

There are [six criteria for interoperability and connectivity](#):

1. Categorisation depending on the sensitivity of the data
 - a. Cat 1: the app does not process any personal data as defined in the GDPR,
 - b. Cat 2: the app processes personal data as defined in the GDPR but does not process sensitive personal data as defined in Article 9 GDPR,
 - c. Cat 3: the app processes sensitive personal data as defined in Article 9 GDPR. Most apps displayed on the mHealth website process sensitive data.
2. Identification of the person in need of care: apps of categories 2 or 3 (referred in criterion 1) uniquely identify the person through his/her Social Security Identification Number at certain moments of the app use.
3. Authentication of the user: the apps of categories 2 or 3 authenticate the app’s user by using a method integrated within the Federal Authentication Service (FAS) or a level equal to or higher than the level determined by the e-Health platform or by an authentication system specific to the provider that meets specific conditions.
4. Verification of the therapeutic relationship: if the use of apps from categories 2 or 3 requires the verification of therapeutic relationship of the app user, these attributes/qualities or relationships are consulted in the relevant authentic sources accessed via the e-Health platform.
5. Interoperability: in case of exchanged structured personal data with information systems of healthcare actors, such as care providers, care institutions or people in need of care, this exchange takes place via open standards, and preferably via the standards set by the e-Health platform, if available.
6. GDPR compliance declaration: the providers of categories 2 or 3 apps confirm that the processing of personal data takes place in accordance with the provisions of the GDPR.

The **M3 level** is related to reimbursement. The Insurance Committee of the NIHDI (National Institute for Health and Disability Insurance, the competent authority for this level), set up a procedure for integrating mobile health in the Belgian reimbursement system. The DMT has to demonstrate a social economic added value. Only apps that have completed levels M1 and M2 of the pyramid can start the integration process in the M3 level. The intention of the NIHDI is not to reimburse mobile apps per se, but their use within the context of a specific care process. An app can be either *temporarily reimbursed (M3-)*, if its socio-economic value is still uncertain, or *structurally reimbursed (M3+)* if its socio-economic value is adequately proven.

Evaluation

The Insurance Committee (the decision-making body) of the NIHDI validated a specific evaluation process for the integration of mobile health apps in the Belgian reimbursement system, which is limited to apps that transfer data remotely.

Manufacturers [notify their app](#) and request an ad hoc expert group to assess their app within a proposed care process (compared to the standard of care).

In their notification, manufacturers must submit a dossier showing

- the clinical and/or socio-economic value of their app: the evidence (whether real-world data or a clinical study) must show that the medical app brings clear value to the specific care process in which it would need to be integrated, i.e. evidence of the added value of the app compared to existing therapeutic alternatives. Evidence can be supported by RCTs, (prospective) studies under routine conditions, international guidelines recommending the technology or application, expert opinions, ...
- the wider benefit to society or savings outside of the healthcare setting in which the cost is generated.

After receipt of a notification by the NIHDI, a specific working group is set up composed of independent and experienced experts in the care process concerned, representatives of the relevant healthcare providers, insurers, patients, and representatives of employers' and employees' organisations.

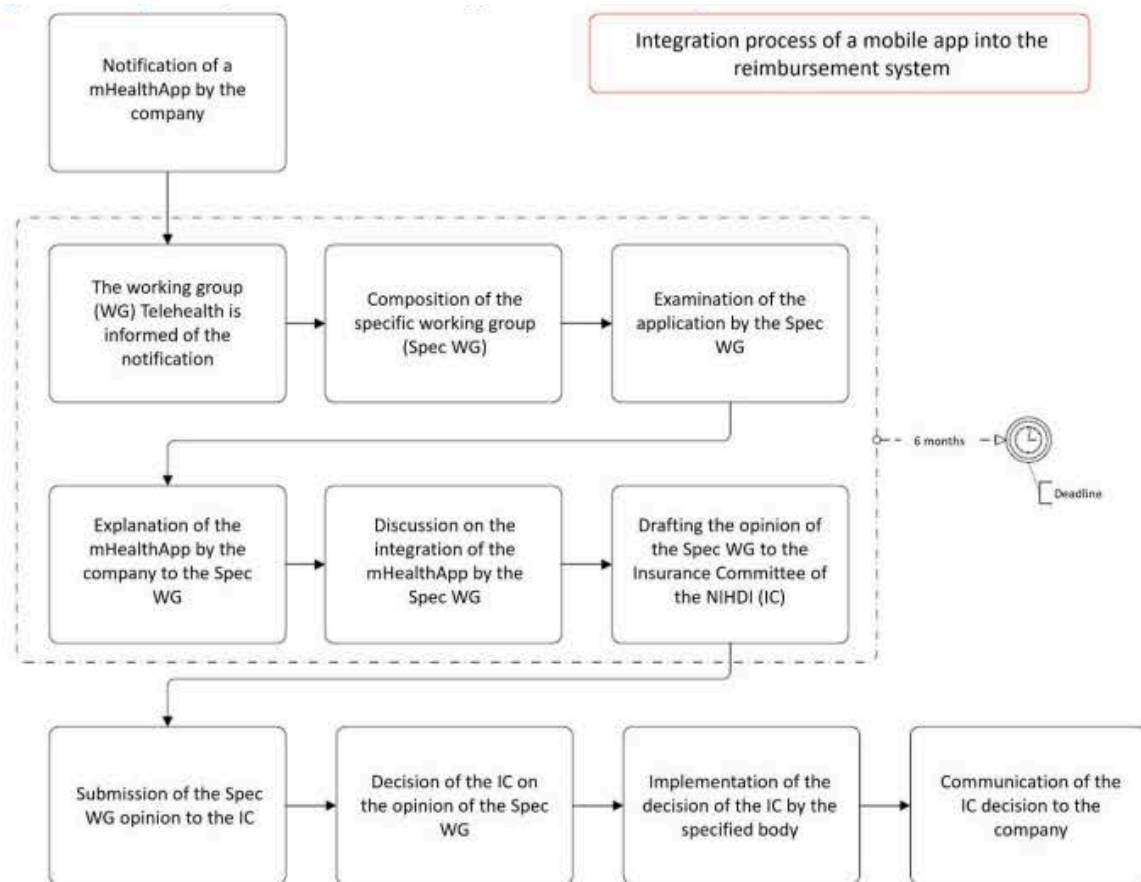
This specific working group examines the extent to which the integration of the mobile app into the healthcare system is interesting and valuable as an addition or replacement to current practice. The company will be invited to present its app and to participate in the working group session. After the meeting, questions are addressed in writing and the dossier is then discussed within the working group. A formal advice (majority of votes) is forwarded to the Insurance Committee of the NIHDI (decision-making body).

The advice of the working group is based on:

- the evaluation of the clinical evidence contained in the dossier;
- the feasibility: the possible integration of the app into the care process;
- the added value: the possible improvement it may bring to current practice;
- the budgetary impact.

The ad hoc working group shall issue its advice within six months (after it is constituted). There is, however, no specific timeline for the Insurance Committee to decide after the latter has received the advice of the ad hoc working group.

At the end of the evaluation process, a final advice is issued by the Insurance Committee of the NIHDI, and the dossier is referred to the most appropriate organ of the NIHDI, competent for the reimbursement mechanism.



Source: [KCE Report 362 \(2023\)](#)

Decision-making (i.e. reimbursement or implementation)

Reimbursement can be temporary or structural and can be implemented via the following existing mechanisms:

- Via the *nomenclature* (list of reimbursed healthcare provisions delivered by healthcare providers). In this case, the healthcare provision is described and could include the use of a CE-marked mobile health app (current scope of the NIHDI).
- Via a “*convention*”, a contract between the NIHDI and healthcare providers or healthcare delivery institutions (a manufacturer never contracts with the NIHDI). This convention system is more flexible than the nomenclature and allows the NIHDI to pay, via a lump sum, for a range of health multidisciplinary care services which may include the use of medical devices and apps. Patients who fulfil the convention’s conditions and who are treated by the institution contracting with the NIHDI are entitled to reimbursement of their follow-up (incl. the app). The competence to sign and fund these rehabilitation contracts was partly transferred to the federated entities in 2014.

In addition to the two above mentioned mechanisms, one cannot exclude, in theory, certain DMTs to be reimbursed via the existing procedure for inscription of medical devices on a *positive (generic or nominal) list*. For medical devices, there are two lists (with different procedures and advisory organs); one for the reimbursement of medical devices (for instance a pacemaker) and one for the reimbursement of non-implantable medical devices (for instance catheters for home urinary self-catheterisation or active bandages). However, the reimbursement of the individual medical device (instead of the reimbursement of a medical device within a care process) is not the approach chosen by the NIHDI for mobile health applications.

In addition to reimbursement via the compulsory health insurance, mobile health apps are currently financed by other means: hospitals can finance them via the innovation budget. If a DMT supports health policies of the federated entities, they can also be reimbursed by the federated entities. Furthermore, sickness funds offer “additional advantages” in exchange for the payment of a subscription fee. Apps can be included in this package. Unlike compulsory insurance services, these services are determined (and can be adapted) by the sickness funds themselves depending on what they consider important for their affiliates.

Sources

- [KCE Reports 362 Evaluation of Digital Medical Technologies](#)
- <https://mhealthbelgium.be/nl/validatiepiramide>

4.5 National Institute for Health and Care Excellence (NICE), United Kingdom

Scope: Digital health technologies

Abbreviations

DHT	Digital Health Technologies
ITP	Innovation and Technology Payment
MIB	MedTech innovation briefings
NHS	National Health Service

Identification

NICE selects digital health technologies that are registered at the NHS Innovation Service (formerly Health Tech Connect). The NHS Innovation Service is made up of a [collective of organisations](#) that all play a part in supporting and assessing innovations along the journey from an idea to implementation and scaling.

The NHS Innovation Service has two key offers for healthcare innovators:

- Information and guidance on regulations, funding, marketing to help you understand and navigate the healthcare system as you develop your innovation
- End-to-end support from different organisations, helping you to take your innovation from idea to adoption faster

When you create an account and register your innovation, you will go through a number of steps and processes so that we can find the right support for you:

- Create an account and complete your innovation record
- Go through our needs assessment process with the service team
- Get the support you need from the organisations that can help you
- Keep your innovation record up to date as you progress

Alongside the Innovation Service is the NICE Medical Technologies Evaluation Programme (MTEP). This identifies and selects medical devices and diagnostic technologies and routes them to appropriate evaluation programmes at NICE. It also develops guidance and advice on the effective and cost efficient use of these technologies for the NHS and its social care partners, and where appropriate, commissions research on the clinical utility of technologies with an underdeveloped evidence base

Selection / eligibility criteria

A DHT that is part of the [Medical Technologies Evaluation Programme](#) is likely to be selected if it meets the [eligibility criteria](#) and:

- it offers substantial benefits to patients, or the health and care system compared with current practice
- the benefits are plausible, clearly described, easily understood, and supported by evidence
- developing guidance would mean faster and more consistent adoption of the technology.

Eligibility criteria

1. The technology must be suitable for development under one of our guidance programmes, and not currently in development.
2. It must be a new technology, or an innovative modification of an existing technology, which will have benefits to the patient or the NHS when judged against comparators.
3. The technology must have one of the following CE marks (or equivalent regulatory approval) or be expected to get one within 12 months:
 - a. a medical device (under EC directive 2007/47/EC or 93/42/EEC)
 - b. an active medical device (under EC directive 90/385/EEC)
 - c. an active implantable medical device (under EC directive 90/385/EEC)
 - d. an in vitro diagnostic medical device (under EC directive 98/79/EC).

The technology must also be available to the NHS, or have plans to be launched in the NHS.

The CE mark is a mandatory conformity mark for all products placed on the single market in the European Economic Area, including medical devices. The mark certifies that a product has met EU consumer safety, health or environmental requirements.

Evaluation

Developers of DHTs should use the following frameworks:

- [The Digital Technology Assessment Criteria \(DTAC\)](#)
- [The Evidence Standards Framework \(ESF\) for digital health technologies](#) (not mandatory)

NICE advice is produced by a faster process that does not include a recommendation on NHS use. Instead, NICE advice (such as [MedTech innovation briefings](#); MIBs) includes

- a description of a regulated medical device or in vitro diagnostics
- its place in the care pathway
- a critical appraisal of the available evidence and some expert opinion.

Decision-making (i.e. reimbursement or implementation)

Whilst national funding (through the Innovation and Technology Payment (ITP)) is no longer provided, the [NHS Accelerated Access Collaborative](#) is encouraging NHS trusts to sustain these innovations through local funding agreements.

[Recommendations](#) provided by the Accelerated Access Collaborative (AAC):

- To be recommended for NHS funding we think DHTs must be authorised by MHRA, pass DTAC and be recommended by NICE.
- Geographical variation is common, as funding, and funding decisions, are still mostly made at a local or regional level by local commissioning bodies (formerly Clinical Commissioning Groups (CCGs), now Integrated Care Systems (ICS))

Additional information & support

NICE evaluates digital health technologies through 4 guidance programmes:

- Technology appraisals guidance.
- Highly specialised technologies guidance.
- Medical technologies guidance.
- Diagnostics guidance.

The programme we use depends on the technology and its expected impact on health and social care spend. The [Office for Digital Health](#) can advise you about our programmes

[Office for Digital Health](#) strives to accelerate our efforts to deliver innovation to the health and care system. We work with partners to identify topics, develop evidence standards and support digital health policy development.

[Scientific advice service](#) offers support for developers of devices, diagnostics and digital technologies. We can help you to develop evidence that shows the value of your technology. And we can provide feedback on your plans.

[Multi-agency advisory service](#) is involved in a new advisory service for developers and adopters of artificial intelligence (AI) and other data driven technologies in England.

[Medicines & Healthcare products Regulatory Agency \(MHRA\)](#) are responsible for regulating medical devices. They will be able to offer advice on submissions to their organisation.

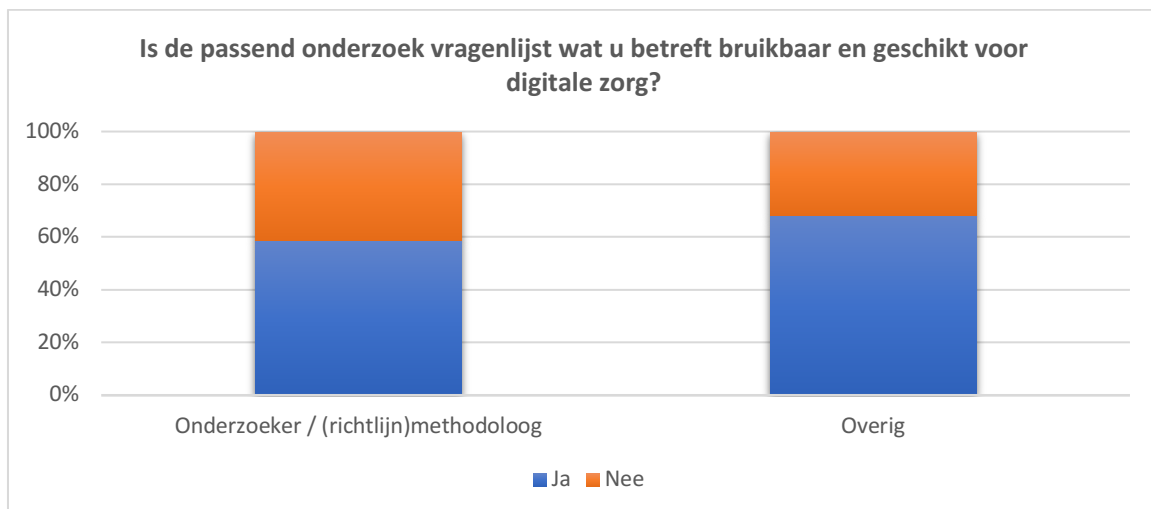
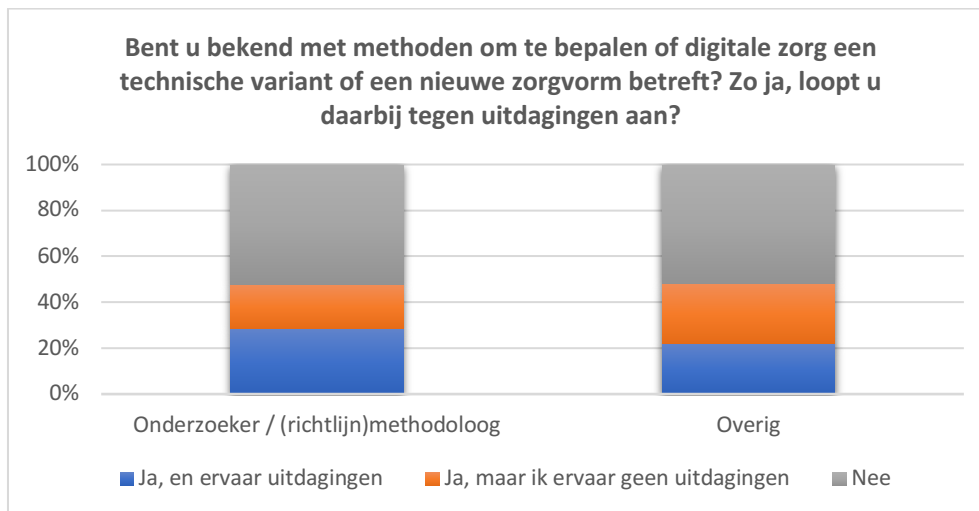
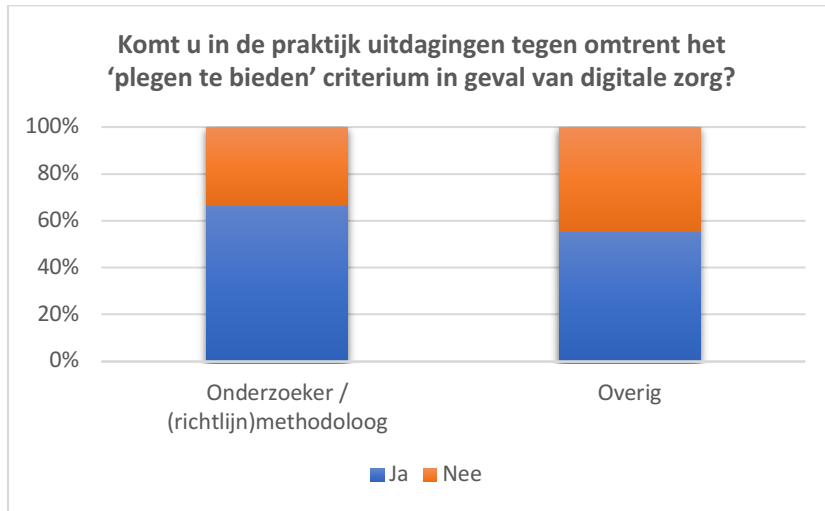
[Care Quality Commission \(CQC\)](#) are responsible for regulating services. They will be able to offer guidance on registration with their organisation.

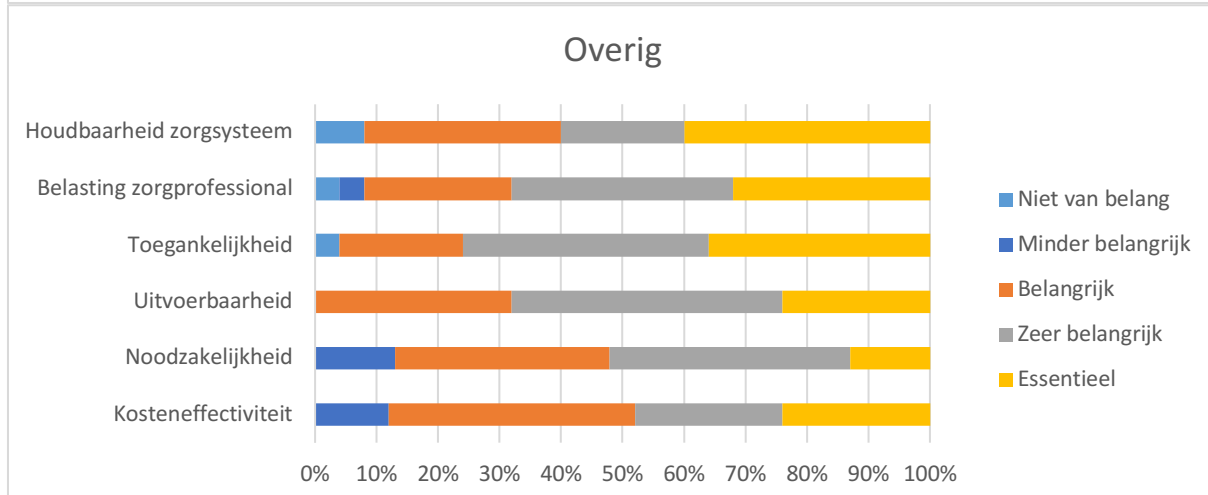
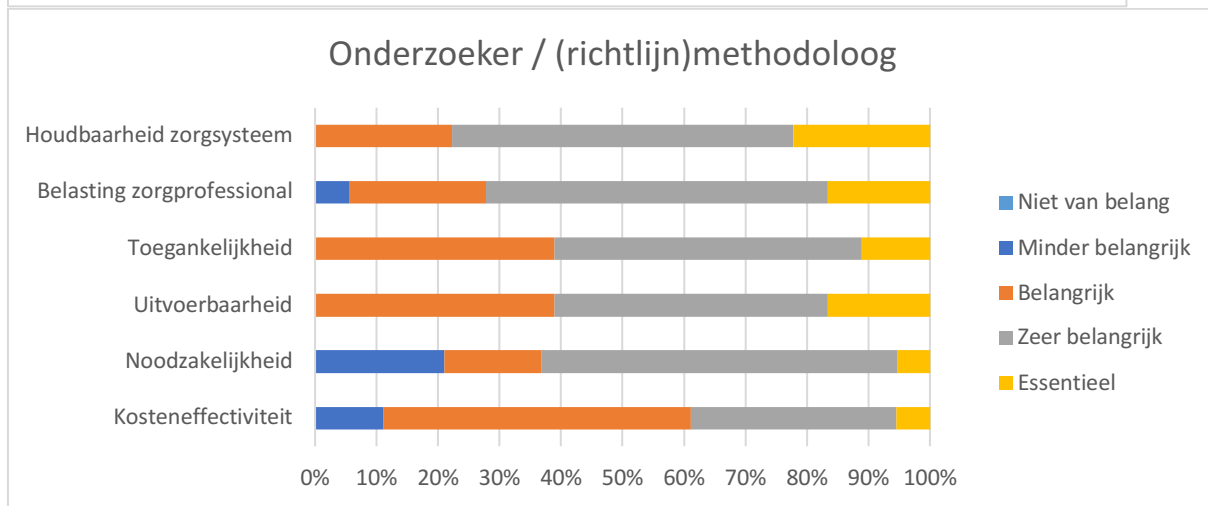
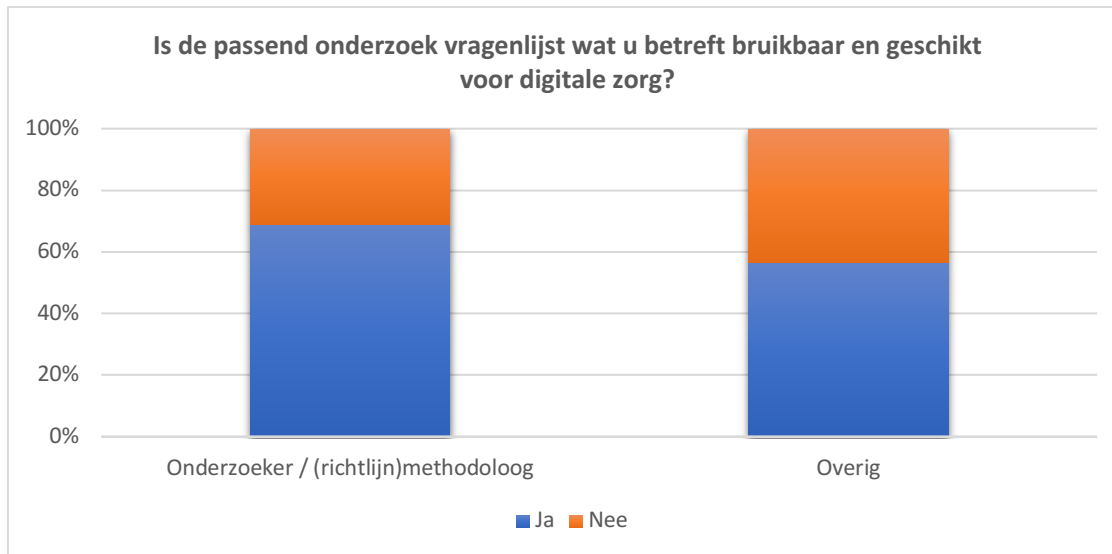
[The Innovative Devices Access Pathway \(IDAP\)](#) is a joint project between NICE, the Medicines and Healthcare products Regulatory Agency (MHRA), HTW and the Scottish Health Technologies Group (SHTG). The aim of the project is to develop a pathway that allows technology developers to provide their products to care professionals and service users at the earliest and safest opportunity.

Source

[Evidence standards framework \(ESF\) for digital health technologies](#)

Bijlage 5. Kwantitatieve analyse vragenlijst: subgroep o.b.v. affiliatie





Bijlage 6. Kwalitatieve analyse vragenlijst

Overstijgende opmerkingen

Algemeen

- Effectiviteit vervangen door werkzaamheid
- Veel innovaties gaan verloren wanneer een RCT wordt vereist
- Verzekeraars hebben weinig behoefte aan valide vergelijkend onderzoek; digitale zorg wordt (mede o.i.v. het IZA) voorgesteld om het acute zorginfarct dicht te lopen
- *Digitale transformatie gaat om het oplossen van (toekomstige en nijpende) maatschappelijke problemen. De trial is in mijn ogen slechts in beperkte mate een instrument die in die queeste iets toevoegt.*
- Digitale zorg niet aan de markt overlaten; noodzaak tot standaardisatie
- Extra bijkomende kosten bij implementatie

Technische variant

Welke uitdagingen liggen er bij de bepaling of specifieke digitale zorg een technische variant betreft?

Algemeen

- Geen methoden voor bepaling / instructie GRADE indirectness
- Geen duidelijke grens
- Onvoldoende voorbeelden (n=2)
- Consensus bereiken lastig
- Aanname vergelijkbare effectiviteit vaak niet plausibel
- Begrip 'technische variant' niet herkend door veldpartijen
- Open pakket veel zorg die niet bewezen effectief is; waarom bij digitale zorg dan wel?

Plegen te bieden

In welke situaties kan het 'plegen te bieden' criterium een punt van discussie vormen?

Wanneer er:

- Geen ruimte is om zorg erbij te krijgen
- Geen ruimte is om zorg los te laten
- Wanneer er verplaatsing is tussen 1e, 2e, 3e lijn
- Onbepaalde bereikbaarheid wordt verwacht van zorgprofessionals
- Nog geen standpunt door beroepsgroep is ingenomen
- Onduidelijkheid is wat beroepsgroepen tot hun 'arsenaal' rekenen

Typen digitale zorg

- Telemonitoring (rol van zorgprofessional bij externe monitoring)
- Zelf thuis monitoring (preventie, diagnostiek, prognostiek)
- Digitale groepstherapie
- Digitale therapie door professional
- Digitaal ondersteunen en taakverdeling naar naasten
- Digitale triage
- Preventieve diagnostiek (e.g. whole-body scan)

RCT wenselijk / haalbaar

Barriers wenselijkheid

- Lage risico's innovatie (n=4)
- Ethische bezwaren (n=4)
- Maatschappelijke bezwaren
- Urgentie bezwaren (n=2)
- Minimale aanpassingen (e.g. software-updates) (n=2)
- Overtuigend bewijs via indirect (observationeel) onderzoek te verkrijgen (n=4)
- Technologie verouderd voor afronding RCT (n=14)
- Context bepaalt effectiviteit van innovatie (n=5)
- Afname motivatie stakeholders bij langdurige RCT
- Extra druk op personeel
- Heterogeniteit van studiepopulatie is juist wenselijk (waar traditionele RCTs veelal een homogene populatie includeren) (n=2)
- Zorginstellingen derven inkomsten als innovaties pas na een trial kunnen worden geïmplementeerd
- Gecontroleerde inzet van innovatie
- Geen ruimte om gaandeweg aanpassingen in design, dosis, frequentie en implementatie aan te brengen (n=2)
- Strak geprotocolleerd METC goedgekeurde traditionele RCT geeft geen reëel en valide beeld en creëert schijnveiligheid met betrekking tot veiligheid effectiviteit en doelmatigheid

Facilitators wenselijkheid

- Communicatie over verantwoording keuzes in design RCT (n=2)
- Controleerde inzet in de praktijk (i.c.m. monitoring)
- Voldoende linked evidence
- Feasibility studie uitgevoerd
- Integraal behandeltraject beoordelen

Barriers haalbaarheid

- Groot onderzoek voor technische variant
- Praktische bezwaren, zoals lange FU tijd (n=4)
- Financiële mogelijkheden startups (n=6)
- Lage prevalentie van ziekte
- Niet geschikt voor adaptieve interventies
- Randomiseren binnen afdeling niet haalbaar
- Vrijheid van keuze patiënten maakt randomiseren naar controlegroep lastig
- Tijdens de trial starten andere ziekenhuizen parallel al met de innovatie

Facilitators haalbaarheid

- Regionale/landelijke multi-stakeholder samenwerkingen en co-creatie (n=3)
- Gegevensuitwisseling
- Gezamenlijke systematiek voor bekostiging (n=2)
- Subsidies voor startups (n=3)
- Platform design
- Stepped-wedge design
- Non-inferiority design
- Pragmatische trials (n=3)
- Trials met interim assessments en korte doorlooptijd (<18 maanden)
- Digitale zorg nog niet publiekelijk beschikbaar
- RCT meer ondersteuning en minder vrijblijvend
- Digitale zorg pas beschikbaar stellen na een RCT
- Analyses relatief simpel vergeleken met RWE
- RCTs makkelijker, goedkoper, en minder tijdrovend maken

Waarom RCT wel wenselijk?

- Digitale zorg is hetzelfde als traditionele zorg (n=4)
- Praktische bezwaren geen argument (n=3)
- Financiële bezwaren geen argument (n=2)
- Kwaliteit zorg veilig en betaalbaar houden (n=2)
- Onnodige zorg voorkomen
- Innovator moet investeren om hun product te kunnen vermarkten

Waarom RCT wel haalbaar?

- Voldoende ruimte aanpassing m.b.v. passend onderzoek vragenlijst

Welke typen wel RCT?

- Behandeling met digitale zorg (n=6)
- AI zorg
- Innovaties met een hoog risico voor de patientveiligheid (n=5)
- Innovaties met potentieel veel arbeidsinzet(besparing), veel kosten(besparing), veel kwaliteits(winst)
- Digitale zorg welke gezondheidsrisico's / kans op fouten in handelen met zich meebrengt
- Substitutie (n=2)
- Communicatie, inclusief bewerken/verwerken/interpreteren door digitale innovatie
- Proof of concept innovaties (effectiviteit en veiligheid)

Welke typen geen RCT?

- Alle digitale zorg leent zich niet voor RCT (n=2)
- Urgente zorg
- Verplaatsing zorg
- Efficiëntie toename
- Wanneer klinisch handelen niet veranderd
- Indicatieuitbreidingen binnen eenzelfde innovatie
- Innovaties waarbij de behandelrelatie (die over een lange periode is opgebouwd) voor een groot deel de impact bepaalt
- Software welke niet als medisch hulpmiddel dient
- Monitoring van patiënten (n=3)
- Add-on
- Communicatiemiddelen, zonder bewerken/verwerken/interpreteren door digitale innovatie (n=3)

Zijn er bepaalde type digitale zorg die de haalbaarheid / wenselijkheid van een RCT beïnvloeden?

- Geen onderscheid maken in verschillende typen digitale zorg (n=4)
- Plaats in het zorgpad (Informatievoorziening, contactvoorbereiding, toegang tot zorg, triage, anamnese, diagnostiek, evaluatie/beleid, controle, evaluatie van zorgproces)
- Complexiteit van de toepassing
- Mogelijkheden tot een controlegroep
- Verschillende doelgroepen en settings
- Toegankelijkheid tot interventies (technologisch en qua vaardigheden)
- Periodieke monitoring
- Continue monitoring
- Risico afhankelijk
- Context afhankelijk
- Risk/Benefit analyse o.b.v. clinical evaluation (onderdeel van CE keuring) bepaalt of RCT nodig is
- Eenvoud van de digitale zorg
- Mate waarin digitale zorg reeds geïmplementeerd is
- O.b.v. verwachte impact op gezondheidsuitkomsten, kosten en benodigd zorgpersoneel
- O.b.v. impact op kosten/arbeid/kwaliteit en ook duurzaamheid voor uitvoering RCT

Observationele / RW evidence

Hoe wordt aangekeken tegen het gebruik van observationele / RW evidence voor de beoordeling van effectiviteit van digitale zorg?

Algemene opmerkingen

- Verplicht dataverzameling evaluatie en bijsturing
- Stel een goedgekeurd protocol op voor monitoring zodat niet 1 individu of organisatie aansprakelijk kan worden gesteld
- RWE = observationeel onderzoek (niet zien als iets compleet anders als trial data) (n=2)
- RWE als term niet geschikt (klinische studies ook 'real-world')
- RWE is valide alternatief op RCT (n=8)
- RWE is valide alternatief op RCT wanneer de innovator kan aantonen dat deze niet haalbaar is
- RWE is aanvulling op andere vormen van bewijs (n=9)
- Goedgekeurde voorwaarden / protocol voor implementatie op voorhand opstellen (n=3)
- Veiligheid op voorhand testen (n=2)
- Gebruik RWE risico- en doelgericht inrichten en toepassen (n=6)
- Bestaand bewijs uiteenzetten, en monitoring inrichten op waar bewijs ontbreekt of schaar is
- RWE gebruik mogelijk afhankelijk van risicoprofiel innovatie (n=2)
- RWE gebruik mogelijk afhankelijk van complexiteit innovatie
- RWE gebruik mogelijk afhankelijk van context / impact van innovatie
- Privacy minder belangrijk
- GDPR / AVG in acht nemen bij continue monitoring
- Continue dataverzameling alleen wanneer de baten in verhouding staan tot de inspanning
- Continue dataverzameling verplicht onder EU-MDR
- RWE alleen als voorlopige toelating, altijd controle of effecten in de praktijk worden behaald
- Uniformeren van data registratie (n=2)
- Hoge kwaliteit databases noodzakelijk
- Complexiteit van analyses is op zichzelf geen reden om het niet te doen
- Uitwerken wat er gebeurt bij een negatieve oordeel bij cyclische beoordeling innovatie (vergoeding intrekken?)
- RWE gebruik niet anders dan bij andere interventies
- RWE monitoring op procesuitkomsten naast effectiviteit (n=2)
- Niet wachten tot alle randvoorwaarden perfect zijn; actie-lerend de (data)infrastructuur verbeteren
- Hoe lager de vooraf geaccepteerde bewijslast, hoe belangrijker de de postmarkering surveillance
- Maak gebruik van ervaringen, methoden, en toepassingen van RWE gebruik in het buitenland
- Onafhankelijke uitvoering / analyse RWE (n=3)
- Standaardisatie noodzakelijk
- Stel post-marketing surveillance / monitoring verplicht (n=3)
- Risico: digitale zorg kan potentieel langs andere lat worden gelegd dan niet-digitale zorg (n=2)
- Vaak alleen non-inferiority aantonen i.c.m. geen toename calamiteiten of incidenten
- Real world data kán veel opleveren, mits aan de voorkant goed is nagedacht over registratie, data validatie en governance. (n=4)
- Observationele studie naar niet CE gemarkeerde innovatie kan alleen maar gedaan worden als clinical investigation volgens ISO14155
- Post-marketing surveillance is noodzakelijk, maar heeft vaak te lijden onder onvoldoende middelen en mandaat
- Alle innovaties tijdig evalueren gaat niet lukken met enkel RCTs
- Overheid moet continue data verzameling faciliteren
- RWE om een pilot te doen en potentie te verkennen, RCT om potentiële innovaties verder te onderzoeken
- RWE is prima om trends te observeren, niet nauwkeurig effectiviteit monitoren

Barriers RWE

- RWE start op een lagere level of evidence dan RCTs; lastig om op high evidence uit te komen
- RWE niet op zichzelf voldoende voor effectiviteit aantonen, altijd nog RCT nodig (n=6)
- Digitale zorg laat over het algemeen kleine impact op effectiviteit zien; upgraden binnen GRADE lastig
- Data infrastructuur en een infrastructuur voor onderzoeksondersteuning nodig (n=5)
- Commerciële partijen nemen veelal niet de moeite om additionele data (bijv. gebruikerservaringen, veiligheid, tijdsinvesteringen) te registreren
- Kwaliteit en interpreteerbaarheid 'messy' datasets
- RWE verzamelen binnen EPD uitdagend; rigide systemen die hier niet op zijn ingericht
- Digitale vaardigheden ontbreken; uitdagend om op te lossen
- Data-vervuiling (bijv. in verzekeringsdata)
- Ethische aspecten (patiënten hebben veelal geen toestemming gegeven)
- Cherry picking
- Gebrek aan expertise omtrent methodologie analyse big data (n=5)
- Zeer complexe analyse methoden (n=2)
- Kans op (selectie)bias (n=3)
- Kans op (te) hoge engagement door gemotiveerde patiënten
- Gebrek aan governance en data eigenaarschap
- RWE werkt vooral hypothese genererend, niet voor het beantwoorden van vraagstellingen (bijv. over effectiviteit) (n=2)
- Missende data vaak groter probleem dan bij RCTs (n=2)
- Om digitale vaardigheden en infrastructuur te onderhouden zijn structurele middelen nodig
- Koppeling verschillende data bronnen kan lastig zijn m.b.t. privacy (AVG/GDPR) (n=2)
- Gebrek aan transparantie
- Verhoogde administratie en registratielast bij monitoring

Facilitators RWE

- Patientveiligheid monitoren van reeds geïmplementeerde innovaties (n=2)
- Biedt flexibiliteit om het zorgpad gaandeweg bij te stellen
- RWE kan via GRADE methode ook op high level of evidence uitkomen (mits voldoende kwaliteit) (n=5)
- Datasolidariteit (i.e. openlijk delen van data)
- Koppeling van databases
- Meenemen met niet-klinische parameters (n=4)
- Methodologische ondersteuning bij analyses RWE data
- Kansen voor continue monitoring, niet alleen als kwaliteitscheck, maar ook ter verbetering AI algoritme
- Meer participanten, grotere sample size
- Representatievere meer heterogene patiëntenpopulatie (n=2)
- Data verzameling over langere tijdsperioden mogelijk
- Digitale vaardigheden steeds minder van belang door slimmere gebruikerstools & interfaces
- Haalbaarheid en tevredenheid van zorgprofessionals en gebruikers kunnen direct worden meegenomen
- Gebruik RWE zal tot snellere implementatie en adoptie leiden, wat binnen de huidige toegankelijkheids crisis belangrijk is
- RWE kan sneller worden uitgevoerd, en zorg sneller geaccepteerd worden (n=3)
- RWE geeft de real-world effecten beter weer dan RCT
- RWE kan effecten post-marketing goed monitoren (n=3)
- Geeft innovator meer ruimte om innovatie aan te passen en te experimenteren
- Kansen voor laagrisico innovaties
- Infrastructuur data registratie huisartsen (COH)

Overige designs

- Kortcyclisch actieonderzoek voor reeds geïmplementeerde innovaties (n=5)
- Studie naar toegevoegde waarde van een grootschaligere studie (e.g. early HTA)
- Quasi experimentele studie (n=2)
- Interrupted time series
- Value of information assessment deel uit laten maken van passend onderzoek vragenlijst
- RWE combineren met natuurlijke experimenten
- Veiligheid en kwaliteitsstudies, vervolgens goede registraties en post-marketing surveillance

Passend onderzoek vragenlijst

Is de passend onderzoek vragenlijst bruikbaar (i.e. valide) en compleet voor digitale zorg?

Bruikbaar (i.e. welke vragen aanpassen?)

- Natuurlijk beloop op pathofysiologie niet relevant voor digitale zorg (zoals telemonitoring) (n=3)
- Is er sprake van volledig nieuwe technologie?
- Specifiekere duiding (bijv. Wat is een klein indicatiegebied?)
- Vragenlijst is gericht op een aandoening, niet de vorm waarin zorg wordt geboden
- Niet geschikt voor digitale ondersteuning (alleen behandeling) (n=3)
- Vragenlijst zegt niets over de wenselijkheid van vormen van onderzoek
- Het kader is de smal, voor klinisch georiënteerd, de overige overwegingen dienen ook mee te worden genomen (n=9)
- Pakket- en validiteitsvragen risicoafhankelijk maken
- Vragenlijst afhankelijk maken van of het een medisch hulpmiddel betreft

Compleet (i.e. welke vragen toevoegen?)

- Wat zijn de kosten en opbrengsten? (n=2)
- Wat is de impact van digitale vaardigheden van de gebruiker op de effectiviteit?
- In hoeverre laat de innovatie ruimte voor voldoende gelijkheid, tijdigheid, patiëntveiligheid, efficiëntie en patientgerichtheid?
- In welke mate wordt het zorgproces verstoord? (n=2)
- Is er een gouden standaard?
- Is er sprake van klinische noodzaak?
- Is er een businesscase te maken (kosten reductie & klinische verbetering)?
- Contextuele factoren (zorgprocessen, passend aan beleid) (n=3)
- Wat is de beschikbaarheid van de innovatie?
- Wat is de bruikbaarheid van de innovatie?
- Wat is de noodzaak tot transformatie?
- Wat zijn de risico's rondom privacy?
- Wat is de ondersteuning bij technische implementatie?
- Geen vragen over mate van generaliseerbaarheid
- Geen vragen over 'feasibility'

Algemene opmerkingen

- Niet anders voor digitale zorg dan voor traditionele zorg (n=3)
- De huidige passend onderzoek vragenlijst heeft de afgelopen jaren niet bij iedere stakeholder tot het gedrag geleid wat nodig is om tot adoptie en opschaling te komen
- Waak voor het afwaarderen van valide onderzoek naar effectiviteit

Overige overwegingen

Welke overige overwegingen zijn belangrijk om mee te nemen bij beoordeling van effectiviteit van digitale zorg?

Kosteneffectiviteit

- Kostenbesparing
- Doelmatigheid
- Productiviteitswinsten
- Budgetimpact
- Investeringskosten meenemen
- PROMS meenemen

Haalbaarheid

- Acceptatie
- Usability,
- Integratie
- Reikwijdte
- Praktische haalbaarheid
- Implementatie
- Voorlopige werkzaamheid

Noodzakelijkheid

- Wie moet het noodzakelijk achten?

Toegankelijkheid

Belasting zorgprofessional

- Niet alleen zorgprofessional
- Mantelzorgers
- Digitale zorg kan drempel verlagen en druk doen toenemen
- Schaalbaarheid

Houdbaarheid systeem

- Overlap met haalbaarheid en kosteneffectiviteit

Additionele overwegingen

- Patiënttevredenheid (n=2)
- Patientvoorkeuren (n=2)
- Duurzaamheid (n=10)
- Implementatie & adoptie innovatie
- Equity (n=4)
- Aanvaardbaarheid
- Arbeidsinzet (n=4)
- Tijdsinvestering (n=2)
- Werkplezier (n=5)
- Betrouwbaarheid / kwaliteit (n=4)
- Snelheid van updatemogelijkheden
- Betaalbaarheid (n=2)
- Substitutie over 1e, 2e , 3e lijn
- Governance en businessmodellen
- Uitvoerbaarheid
- Toekomstbestendigheid
- Vereiste bewijskracht effectiviteit verdisconteren met impact op andere overwegingen
- Impact op therapeutische behandel relatie
- Belasting patient
- Gebruiksvriendelijkheid
- Uitlegbaarheid
- Impact op toe- of afname complexiteit zorgsysteem

Overige opmerkingen

- Er wordt verwezen naar ons rapport over het 'Plegen te bieden-criterium'. Goed om te weten dat in PIP-3 in bijlage 2 een soort addendum hierop is gekomen. <https://www.zorginstituutnederland.nl/publicaties/rapport/2013/10/18/pakketbeheer-in-de-praktijk-deel-3>. Ook is er inmiddels een PIP-4, aan de hand van Jacqueline Zwaap, is zij ook betrokken? En is JZ ook meegenomen?
- Voor andere onderzoeksdesigns, zie ook Zorgevaluatie en Gepast Gebruik (ZE&GG).
- Softwareupdates mogen in het kader van de MDR worden doorgevoerd
- Zie ook recent rapport van Rathenau instituut over innovatie die opgave gericht moet zijn (en niet per se het omzetten van het oude in een digitale variant)
- Allereerst moet er eenduidigheid komen in de te gebruiken begrippen, zodat de discussie over (kosten-)effectiviteit niet verstoord wordt door verwarring over termen. Vanuit het NHG is een begrippenkader ontwikkeld dat hiervoor als basis kan dienen (zie <https://www.nhg.org/praktijkvoering/digitale-zorg/begrippenkader-digitale-zorg/>).
- Vervolgens is het gewenst om de verschillende vormen van digitale zorg te koppelen aan het zorgproces. Voor de huisartsenzorg hebben we dit uitgewerkt in een figuur met de plaats van digitale toepassingen in het zorgproces (zie <https://www.nhg.org/thema/digitale-zorg/visie-op-digitale-huisartsenzorg/digitalisering-e-health-digitale-huisartsenzorg-wat-is-het-verschil/>).
- Kijk vooral ook naar factoren waar nu ook andere organisaties op wegen zoals het 'Transformatiemodel' van het kenniscentrum Digitale Zorg, het 'Honinggraatmodel' van Vilans, 'Digital Health Assessment Technologie-database' van Orcha, 'Validatie-aanpak' van het NeLL, etc etc
- bij AI-toepassingen gelden extra criteria bij de overige overwegingen (zie [https://www.europarl.europa.eu/thinktank/en/document/EPRS_STU\(2022\)729512](https://www.europarl.europa.eu/thinktank/en/document/EPRS_STU(2022)729512))
- Ik zou voor een meer dynamische vorm van experimenten en validaties kiezen, gebruik makend van de technieken van decision under uncertainty. http://dx.doi.org/10.1007/978-3-030-10531-0_7
- Schünemann HJ, Cuello C, Akl EA, et al; GRADE Working Group. GRADE guidelines: 18. How ROBINS-I and other tools to assess risk of bias in nonrandomized studies should be used to rate the certainty of a body of evidence. J Clin Epidemiol. 2019 Jul;111:105-114. doi: 10.1016/j.jclinepi.2018.01.012. Epub 2018 Feb 9. PMID: 29432858; PMCID: PMC6692166.
- Cuello CA, Morgan RL, Brozek J, Verbeek J, Thayer K, Ansari MT, Guyatt G, Schünemann HJ. Case studies to explore the optimal use of randomized and nonrandomized studies in evidence syntheses that use GRADE. J Clin Epidemiol. 2022 Dec;152:56-69. doi: 10.1016/j.jclinepi.2022.09.014. Epub 2022 Oct 2. PMID: 36198367.
- Schünemann HJ, Tugwell P, Reeves BC, Akl EA, Santesso N, Spencer FA, Shea B, Wells G, Helfand M. Non-randomized studies as a source of complementary, sequential or replacement evidence for randomized controlled trials in systematic reviews on the effects of interventions. Res Synth Methods. 2013 Mar;4(1):49-62. doi: 10.1002/jrsm.1078. Erratum in: Res Synth Methods. 2013 Sep;4(3):287-9. PMID: 26053539.
- Het start met het uniformeren van dataregistratie (bijv. zo min mogelijk open tekst), dat is de basis om goed onderzoek te kunnen doen. Zie ook <https://www.neurotransdata.com/en/> waar het wel gelukt is om data gestructureerd te verzamelen en dit er uiteindelijk toe heeft geleid dat ze een platform met meerdere tools (waaronder voorspelmodel) hebben kunnen ontwikkelen: <https://www.neurotransdata.com/en/destiny>
- Kansen liggen m.i. op het gebied van counterfactual reasoning (zie ook voorbeelden in boek van Judea Pearl: *The Book of Why*). Er is veel literatuur hierover (bijv. <https://www.jstor.org/stable/23891788> en dan 'snowballing')