

**Methodologisch
kader**

Samen beslissen met uitkomstinformatie



Colofon

Dit methodologisch kader is ontwikkeld door de *Werkgroep Data(on)mogelijkheden* van het Linnean Initiatief in samenwerking met de *Uitkomstgerichte Zorg Werkgroep Samen Beslissen* van het Ministerie van Volksgezondheid, Welzijn & Sport (VWS).

Samenstelling en coördinatie

Auteurs

Martijn A. Oude Voshaar & Margrietha van der Linde (Linnean Initiatief)

Projectleider

Enny te Gussinklo – Brouns (Linnean Initiatief)

Samenwerkende partij

Werkgroep Uitkomstgerichte Zorg Samen Beslissen (gefaciliteerd door VWS)

Loes Peters (Federatie Medisch Specialisten), Danielle Vermond (Federatie Medisch Specialisten), Annemarie van Leeuwen (Patiëntenfederatie Nederland), Marjolein Boereboom (NVZ), Martine de Jong (ZKN), Anne Vogelaar (Santeon), Ester Rake (Santeon), Ciska Pruijssers – Breas (UMCNL), Gera Welker (UMCNL), Wendy Enthoven (ICTU), Thom Spruijt (ICTU)

Met dank aan

Methodologisch experts

Michiel A.J. Luijten (Department of Epidemiology and Data Science, Amsterdam UMC), Eveline J.A. Wiegers (Department of Neurology, Erasmus MC), Carolien C.H.M. Maas (Department of Clinical Epidemiology, LUMC), Xi Li (Department of Public Health, Erasmus MC), Stella M. Marceta (Health Technology Assessment, Erasmus School of Health Policy and Management)

Deelnemers Linnean Expertsessies

Willem Jan Bos, Joke Bais, Alice Bakker, Elvan Simsek-Cakar, Annemarie van Leeuwen, Dominique Sprengers, Else Wolak, Domino Determann, Delian Hofmann, Daphne van Deursen

Linnean Werkgroep Data(on)mogelijkheden

Klankbordgroep Federatie Medisch Specialisten Samen Beslissen

Redactie en Vormgeving

Eindredactie

Peter van den Boom

Grafische Vormgeving

Talitha Hoos

Inhoud

Inhoud	3
Begrippenlijst	5
1. Inleiding	10
2. Definities	12
2.1 Samen beslissen	12
2.2 Uitkomstinformatie	13
2.3 Samen beslissen met uitkomstinformatie	14
3. Hoe integreer je uitkomstinformatie bij Samen Beslissen?	15
Inleiding	15
Randvoorwaarden	17
Randvoorwaarden op meso- en macroniveau	17
Grote en kleine beslissingen	19
Gezondheidsvaardigheden en inclusiviteit	19
Stap 1: Keuze	20
3.1 (H)erkenning van het keuzemoment	20
Stap 2: Opties	21
3.2 Beschikbare behandelopties	21
3.3 Relevante aspecten voor de patiënt	22
3.4 Vergelijking van opties op relevante aspecten	24
Stap 3: Voorkeur	26
3.5 Absolute grenzen en dealbreakers	26
Stap 4: Beslissing	27
3.6 Gezamenlijke afweging	27
3.7 Congruentie met waarden en voorkeuren	29
Referenties	31

4. Hoe verzamel en duid je uitkomstinformatie: N=1	32
Inleiding	32
Praktische randvoorwaarden voor N=1-terugkoppeling in de spreekkamer	32
4.1 Relevantie van de uitkomstinformatie	33
4.2 Meetmomenten en timing	34
4.3 Validiteit van individuele metingen	35
4.4 Betrouwbaarheid van individuele metingen	37
4.5 Responsiviteit: vermogen om verandering te meten	38
4.6 Meetfout en betrouwbaarheid van verandering	39
4.7 Interpretatie van ernst, drempelwaarden en relevante verandering	41
4.8 Begrijpelijke en doelgerichte presentatie van uitkomstinformatie	43
Referenties	45
5. Hoe verzamel en duid je uitkomstinformatie: 'patients like me'	47
Inleiding	47
5.1 Prognose voor de patiënt	50
5.1.1 Relevantie van de voorspelde uitkomst	50
5.1.2 Kenmerken van de populatie	50
5.1.3 Type en kwaliteit van de gebruikte data	52
5.1.4 Consistentie van meting	53
5.1.5 Externe validatie	54
5.1.6 Onzekerheid rond de voorspelling	55
5.1.7 Discriminatievermogen	56
5.1.8 Kalibratie	57
5.1.9 Klinische bruikbaarheid	58
5.2 Keuze tussen behandelingen	59
5.2.1 Causaliteit van behandel-effecten	59
5.2.2 Onzekerheid bij behandelvergelijkingen	62
5.2.3 Heterogeniteit van behandel-effecten	63
5.2.4 Begrijpelijke presentatie van behandelvergelijkingen	64
Referenties	68

Begrippenlijst

Average Treatment Effect (ATE, gemiddeld behandel-effect): Het gemiddelde verschil in uitkomst tussen een behandelgroep en een controle- of vergelijkingsgroep over de gehele onderzochte populatie. Zegt niets over wat het effect is voor een individuele patiënt.

Area Under the ROC Curve (AUC, ook: C-statistiek): Een maat voor het discriminatievermogen van een voorspelmodel bij binaire uitkomsten. Waarden liggen tussen 0,5 (niet beter dan toeval) en 1,0 (perfecte discriminatie). Een waarde $\geq 0,70$ wordt doorgaans als goed beschouwd.

Baselinemeting: De eerste meting die plaatsvindt aan het begin van een zorgtraject of behandeling, en die als referentiepunt dient voor alle latere metingen. Zonder baselinemeting is het verloop van een uitkomst over tijd moeilijk te interpreteren.

Betrouwbaarheidscoefficiënt (o.a. Cronbach's alfa): Een statistische maat die aangeeft in hoeverre de items van een vragenlijst samen consistent hetzelfde onderliggende construct meten (interne consistentie). Voor individueel gebruik wordt een waarde van circa 0,90 of hoger aanbevolen.

Betrouwbaarheidsinterval: Een interval dat aangeeft binnen welk bereik de werkelijke waarde of uitkomst met een bepaalde zekerheid (bijv. 95%) waarschijnlijk ligt. Een breed interval duidt op meer onzekerheid; een smal interval op een preciezere schatting.

Causaliteit / causale inferentie: De mate van zekerheid waarmee een gevonden verschil in uitkomst daadwerkelijk wordt veroorzaakt door de behandeling, en niet door andere (verstorende) factoren. Vereist specifieke aannames en studiedesigns om te kunnen vaststellen.

Conditional exchangeability: Causale aanname die stelt dat, nadat gecorrigeerd is voor alle relevante patiëntkenmerken, de toewijzing aan een behandeling onafhankelijk is van de potentiële uitkomsten. In de praktijk niet verifieerbaar uit de data zelf.

Confounding (vertekening): Het verschijnsel waarbij een variabele (de confounder) zowel de behandelkeuze als de uitkomst beïnvloedt, waardoor een gevonden verband tussen behandeling en uitkomst vertekend kan zijn. Bijvoorbeeld roken bij het effect van gele vingers op overlijden. Dit is vooral een risico bij observationele data.

Consistency (causaliteitsaanname): De aanname dat de geobserveerde uitkomst bij een patiënt die een behandeling ontvangt overeenkomt met de potentiële uitkomst onder diezelfde behandeling. Vereist dat de behandeling eenduidig is gedefinieerd en op vergelijkbare wijze is toegepast.

Directed Acyclic Graph (DAG): Een conceptueel diagram waarin pijlen de veronderstelde causale relaties tussen variabelen weergeven. Wordt gebruikt om inzichtelijk te maken welke variabelen als confounders moeten worden meegenomen in een analyse.

Decision Curve Analysis (DCA): Een methode om de klinische bruikbaarheid van een voorspelmodel te evalueren door het netto klinische voordeel te berekenen over een reeks drempelwaarden. Hiermee kan worden beoordeeld of gebruik van het model leidt tot betere beslissingen dan altijd of nooit behandelen.

Discriminatievermogen (C-statistiek / AUC): Het vermogen van een voorspelmodel om onderscheid te maken tussen patiënten met een gunstige en een minder gunstige uitkomst. Voor binaire uitkomsten uitgedrukt als C-statistiek of AUC; voor continue uitkomsten als R^2 .

Externe validatie: Het testen van een voorspelmodel in een populatie die niet is gebruikt voor de ontwikkeling van het model. Aantoonbaar noodzakelijk om te beoordelen of de voorspellingen ook buiten de ontwikkelpopulatie betrouwbaar blijven. Heterogeneous Treatment Effects (HTE, heterogene behandel-effecten): De variatie in behandelrespons tussen individuele patiënten. Patiënten kunnen sterk verschillend reageren op dezelfde behandeling, afhankelijk van kenmerken zoals leeftijd, ziekte-ernst of comorbiditeit.

Intraclass Correlatiecoëfficiënt (ICC): Een statistische maat voor de overeenstemming tussen en binnen beoordelaars bij uitkomsten die professioneel oordeel vereisen (bijv. klinische scores of beeldvormingsbeoordelingen). Voor individueel gebruik wordt een $ICC \geq 0,90$ aanbevolen.

International Consortium for Health Outcomes Measurement (ICHOM): Internationale organisatie die gestandaardiseerde uitkomstensets ontwikkelt per aandoening, met als doel vergelijking van zorgkwaliteit en patiëntuitkomsten wereldwijd mogelijk te maken.

Immortal time bias: Een vorm van vertekening in observationeel onderzoek die ontstaat wanneer een periode vóór de start van een behandeling ten onrechte als behandelperiode wordt meegeteld. Kan leiden tot overschatting van het gunstige effect van een behandeling.

Kalibratie: De mate waarin de voorspelde kansen van een model overeenkomen met de werkelijk waargenomen uitkomsten. Een goed gekalibreerd model voorspelt bijvoorbeeld dat 20% van de patiënten met een geschatte kans van 0,20 de uitkomst ook daadwerkelijk ervaart.

Kalibreren (model updaten): Het aanpassen van een voorspelmodel aan de lokale populatie of zorgcontext, om te zorgen dat de voorspelde kansen overeenkomen met de werkelijk waargenomen uitkomsten in die populatie.

Keuzehulp: Een hulpmiddel (bijv. folder, website of app) dat patiënten ondersteunt bij het nemen van een geïnformeerde behandelbeslissing door informatie over opties, voor- en nadelen en uitkomsten op een toegankelijke manier te presenteren.

Logistische regressie: Een traditioneel statistisch model dat de kans op een binaire uitkomst (bijv. wel/geen complicatie) voorspelt op basis van een combinatie van patiëntkenmerken.

Meetfout, SEM (Standard Error of Measurement): De variatie die inherent is aan een meetinstrument, los van werkelijke verandering in de toestand van de patiënt. Hoe groter de meetfout, hoe onbetrouwbaarder individuele metingen zijn en hoe groter de minimale detecteerbare verandering.

Minimally (Clinically) Important Change (M(C)IC): De kleinste gemiddelde verandering in een uitkomstmaat die een groep patiënten als klinisch relevant ervaart. Ontwikkeld op groepsniveau; niet geschikt voor interpretatie van verandering bij individuele patiënten (gebruik daarvoor de SDC).

Observationele data: Gegevens die zijn verzameld in de dagelijkse klinische praktijk, zonder experimentele opzet of randomisatie. Weerspiegelen de werkelijkheid goed, maar zijn gevoelig voor confounding en andere vormen van vertekening.

Overfitting: Het verschijnsel waarbij een voorspelmodel te nauw is afgestemd op de specifieke kenmerken van de trainingsdata, inclusief ruis, waardoor het minder goed generaliseert naar nieuwe patiënten of settings.

Plafondeffect: Het verschijnsel waarbij een groot deel van de patiënten al de maximale score op een uitkomstmaat bereikt, waardoor verdere verbetering niet meer meetbaar is. Beperkt de responsiviteit van het instrument.

Positivity (causaliteitsaannname): De aanname dat voor elke combinatie van relevante patiëntkenmerken een reële kans bestaat om elke behandeling te ontvangen. Ontbreekt deze overlap, dan is een geldige causale vergelijking niet mogelijk.

Potential outcomes (potentiële uitkomsten): Theoretisch concept uit de causale literatuur: voor elke patiënt worden twee (of meer) hypothetische uitkomsten gedefinieerd, één per behandeloptie. Omdat slechts één uitkomst werkelijk wordt geobserveerd, blijft de andere altijd onbekend.

PROMs (Patient-Reported Outcome Measures): Gestandaardiseerde vragenlijsten waarmee patiënten zelf hun gezondheid, klachten, functioneren of kwaliteit van leven rapporteren, zonder tussenkomst van een zorgprofessional. Voorbeelden: pijn, vermoeidheid, kwaliteit van leven.

Propensity score: De geschatte kans dat een patiënt, gegeven zijn of haar kenmerken, een bepaalde behandeling ontvangt. Wordt gebruikt om confounding in observationele studies te verminderen door behandelgroepen vergelijkbaarder te maken.

Puntschatting: Een enkelvoudige waarde die de beste schatting weergeeft van een uitkomst of effect (bijv. een gemiddelde, mediaan of kans). Altijd gepaard gaand met onzekerheid, uitgedrukt in een betrouwbaarheids- of predictie-interval.

Randomisatie / gerandomiseerde studie (RCT): Een onderzoeksopzet waarbij deelnemers willekeurig worden toegewezen aan behandelgroepen. Door randomisatie zijn behandelgroepen gemiddeld vergelijkbaar, waardoor gevonden verschillen in uitkomst causaal aan de behandeling kunnen worden toegeschreven.

Recall bias: Vertekening bij patiëntgerapporteerde uitkomsten doordat patiënten bij het terugdenken aan een langere periode recentere ervaringen zwaarder laten wegen. Een kortere of geen recallperiode kan dit beperken.

Recallperiode: De tijdsperiode waarop een patiëntgerapporteerde uitkomstmaat betrekking heeft (bijv. 'de afgelopen week' of 'de afgelopen maand'). Een te lange recallperiode vergroot de kans op recall bias.

Responsiviteit: Het vermogen van een meetinstrument om veranderingen in de gezondheidstoestand te detecteren die klinisch relevant zijn in een bepaald zorgtraject. Vereist dat het instrument voldoende ruimte biedt en geen vloer- of plafondeffecten vertoont.

Smallest Detectable Change (SDC, minimale detecteerbare verandering): De kleinste verandering in een uitkomstmaat die met redelijke zekerheid kan worden onderscheiden van meetfout bij een individuele patiënt. Berekend als $SDC = 1,96 \times \sqrt{2} \times SEM$. De relevante referentie voor individuele veranderingsinterpretatie (niet de MIC/MCID).

Selectiebias: Vertekening die ontstaat doordat de patiënten die in een dataset voorkomen niet representatief zijn voor de bredere populatie waarop conclusies worden getrokken, bijv. door inclusiecriteria of uitval.

Stratificatie: Het indelen van patiënten in subgroepen (strata) op basis van gedeelde kenmerken (bijv. leeftijd, geslacht, diagnose, ziektestadium). Zorgt voor een relevantere vergelijking bij 'patients like me'-toepassingen, maar leidt tot kleinere subgroepen en meer statistische onzekerheid.

Target trial emulation: Een methodologisch raamwerk voor observationeel onderzoek waarbij de opzet van een hypothetische gerandomiseerde studie (target trial) wordt nagebootst met behulp van observationele data. Bedoeld om causale behandeffecten zo goed mogelijk te schatten.

Toevalsvariatie (statistische ruis): Willekeurige schommelingen in meetwaarden die niet het gevolg zijn van werkelijke verandering in de toestand van de patiënt, maar van meetfout of toevallige factoren. Bij kleine subgroepen of lage betrouwbaarheid speelt toevalsvariatie een grotere rol.

Validiteit: De mate waarin een meetinstrument daadwerkelijk meet wat het beoogt te meten. Verschillende vormen van vertekening (bijv. recall bias, disability paradox, confounding) kunnen de validiteit aantasten.

Vloereffect: Het verschijnsel waarbij een groot deel van de patiënten al de minimale score op een uitkomstmaat behaalt, waardoor verdere verslechtering niet meer meetbaar is.

Voorspelmodel: Een statistisch model dat de kans op een bepaalde uitkomst voor een individuele patiënt schat op basis van meerdere kenmerken (voorspellers), afgeleid uit historische patiëntgegevens. Geeft een individuele kans schatting, in tegenstelling tot stratificatie dat werkt met groepsgegevens.

Voorspeller (predictor): Een patiëntkenmerk dat in een voorspelmodel wordt gebruikt om een uitkomst te voorspellen, zoals leeftijd, geslacht, diagnose of een laboratoriumwaarde. Moet beschikbaar zijn op het moment van de (behandel)beslissing.

Zorgprofessionals: Een verzamelterm voor zorgverleners, waaronder artsen, verpleegkundigen, verpleegkundig specialisten, psychologen en fysiotherapeuten, die betrokken zijn bij diagnose, behandeling, begeleiding en gezamenlijke besluitvorming met patiënten.

1. Inleiding

Uitkomstinformatie wordt steeds vaker gebruikt in de spreekkamer om patiënten en zorgprofessionals te ondersteunen bij het nemen van (behandel) beslissingen. Dit biedt veel kansen om zorg beter af te stemmen op wat voor de patiënt belangrijk en mogelijk is. Tegelijkertijd is het van groot belang dat deze informatie op een zorgvuldige en begrijpelijke manier wordt verzameld, gepresenteerd en toegepast. Niet alle cijfers vertellen het hele verhaal, en zonder duidelijke uitleg of context kunnen zorgprofessionals en patiënten verkeerde verwachtingen krijgen of keuzes maken die niet goed passen bij de situatie.

Om uitkomstinformatie te gebruiken bij samen beslissen is het belangrijk om te begrijpen hoe deze informatie op een correcte en verantwoorde manier gebruikt kan worden. Dit methodologisch kader heeft twee functies:

- **De eerste functie** van dit kader is het bieden van besliskundige handvatten om uitkomstinformatie te integreren met waarden en voorkeuren van de patiënt. Dit stelt zorgprofessionals en patiënten in staat om uitkomst-informatie systematisch mee te nemen in een bredere afweging met voor de patiënt belangrijke aspecten. Een voorbeeld is de mogelijkheid om tijdens de behandeling dicht bij familie en naasten te blijven.
- **De tweede functie** van dit kader is het bieden van methodologische standaarden om uitkomstinformatie correct te duiden: wat zeggen de data wél en wat niet? Dit geldt zowel voor N=1-situaties, waarin individuele trends in de tijd centraal staan, als voor 'patients like me'-situaties, waarin uitkomsten worden vergeleken met die van andere patiënten.

De methodologische handvatten zijn geformuleerd als praktische vragen die eindgebruikers en ontwikkelaars zich in de praktijk kunnen stellen. Elke vraag is gekoppeld aan een specifiek methodologisch of besliskundig concept. In combinatie bieden de vragen een samenhangend geheel dat de belangrijkste methodologische en besliskundige principes dekt. Het kader bestaat uit drie hoofdstukken die elk een eigen invalshoek hebben en in verschillende mate aansluiten bij de behoeften van diverse stakeholders:

1. *Hoe integreer je uitkomstinformatie bij samen beslissen?* is vooral relevant voor artsen en patiënten, omdat het ingaat op het integreren van uitkomst-informatie in samen beslissen. Dit kan als basis gebruikt worden voor het ontwikkelen van scholings- en ondersteuningsmateriaal voor het samen beslissen met uitkomstinformatie.

- 
2. *Hoe verzamel en duid je uitkomstinformatie: N=1* richt zich primair op partijen die uitkomstinformatietools ontwikkelen of beschikbaar stellen. Voor zorgprofessionals vormen zij achtergrondinformatie bij het beoordelen en gebruiken van dergelijke tools om samen te beslissen.
 3. *Hoe verzamel en duid je uitkomstinformatie: 'patients like me'* is in de eerste plaats bedoeld voor ontwikkelaars van uitkomstinformatie, zoals kwaliteitsregistraties en informatiespecialisten.

Bij elke vraag verwijzen we naar de relevante methodologische literatuur en internationale standaarden die het concept onderbouwen.

2. Definities

2.1 Samen beslissen

Wij sluiten aan bij de landelijke definitie van Samen beslissen: een proces waarin zorgprofessional en patiënt samen bespreken welke diagnostiek, behandeling of zorg het beste past bij de patiënt, waarbij alle opties, hun voor- en nadelen, en de voorkeuren en omstandigheden van de patiënt worden meegenomen.

In dit proces brengen patiënt en zorgprofessional hun kennis en ervaring samen: de patiënt kent de eigen situatie, wensen en behoeften; de zorgprofessional weet welke behandelingen, onderzoeken, medicatie of ondersteuning medisch mogelijk en passend zijn. Door deze perspectieven te combineren, ontstaat de keuze die het beste aansluit bij het leven van de patiënt.

Het proces wordt doorgaans beschreven in vier stappen:

- **Stap 1 Keuze:** patiënt en zorgprofessional erkennen samen dat er een keuze te maken is, en dat beide perspectieven (van patiënt én zorgprofessional) belangrijk zijn voor het besluit.
- **Stap 2 Opties:** de verschillende behandel- of zorgmogelijkheden worden besproken, inclusief de voor- en nadelen; ook het afzien van behandeling kan een optie zijn.
- **Stap 3 Voorkeur:** de patiënt vertelt over persoonlijke waarden, wensen en omstandigheden, de zorgverlener koppelt deze aan de beschikbare medische mogelijkheden.
- **Stap 4 Beslissing:** patiënt en zorgprofessional nemen gezamenlijk een besluit door medische kennis en persoonlijke voorkeuren t.o.v. de opties samen af te wegen.

2.2 Uitkomstinformatie

- Een uitkomst gaat over het resultaat van een behandeling, zorgtraject of het natuurlijk beloop van een aandoening op de gezondheidstoestand, het welbevinden of het dagelijks functioneren van een patiënt.
- Uitkomstinformatie is systematisch verzamelde informatie over uitkomsten. Dit betekent dat uitkomsten op een vergelijkbare en gestandaardiseerde manier, en op consistente meetmomenten in het zorgproces worden gemeten en vastgelegd. Dit zorgt ervoor dat uitkomsten betrouwbaar kunnen worden geïnterpreteerd en gebruikt in het gesprek tussen patiënt en zorgprofessional.
- Het kan daarbij gaan om klinische uitkomsten, zoals longfunctie of bloedwaarden, of om door patiënten gerapporteerde uitkomsten (PROMs), zoals pijn, vermoeidheid of kwaliteit van leven.
- Uitkomstinformatie kan men zowel op individueel niveau als op groepsniveau verzamelen. De betekenis ervan hangt altijd af van de context waarin zij wordt geïnterpreteerd, zoals de kenmerken van de patiënt, het ziekteverloop en diens persoonlijke voorkeuren en omstandigheden.



2.3 Samen beslissen met uitkomstinformatie

Samen beslissen met uitkomstinformatie betekent dat zorgprofessional en patiënt actief gebruik maken van uitkomsten die relevant zijn voor de doelen en waarden van de patiënt in hun gezamenlijke besluitvorming over behandeling of zorg.

Afhankelijk van de situatie kan dit op twee manieren vorm krijgen. Voor beide toepassingen geldt dat uitkomstinformatie nooit op zichzelf wordt gebruikt. Deze wordt altijd ingebed in een gesprek waarin men de persoonlijke waarden, wensen en omstandigheden van de patiënt meeneemt. Alleen in die bredere context kan uitkomstinformatie bijdragen aan een keuze die niet alleen medisch verantwoord is, maar ook werkelijk past bij het leven van de patiënt.

- **Op basis van eigen uitkomsten van de patiënt (N=1):** uitkomstinformatie over de eigen gezondheidstoestand van de patiënt vormt de basis voor het gesprek. Ze maakt het verloop van de ziekte of aandoening inzichtelijk en bespreekbaar, helpt om voor de patiënt belangrijke onderwerpen aan te kaarten en ondersteunt gezamenlijke besluitvorming over het voortzetten of aanpassen van de behandeling.
- **Op basis van uitkomstinformatie van vergelijkbare patiënten ('patients like me'):** hierbij wordt gebruikgemaakt van geaggregeerde data van vergelijkbare patiënten die een behandeling eerder hebben ondergaan. Deze informatie kan men op twee manieren inzetten:

Ten eerste kan zij patiënten informeren over de te verwachten uitkomsten van een behandeling die wordt overwogen, en de mogelijke voor- en nadelen in termen van gezondheidseffecten.

Ten tweede kan zij helpen bij het vergelijken van uitkomsten van verschillende behandelopties. In beide gevallen ondersteunt de informatie het gesprek tussen patiënt en zorgprofessional over welke optie het beste past bij de situatie en voorkeuren van de patiënt.

3. Hoe integreer je uitkomstinformatie bij Samen beslissen?

Inleiding

Dit hoofdstuk biedt handvatten voor het integreren van uitkomstinformatie in het proces van Samen beslissen. De structuur volgt de vier stappen die landelijk zijn vastgesteld voor Samen beslissen. Daarom wordt basiskennis van Samen beslissen verondersteld. Dit hoofdstuk zal zich specifiek richten op hoe uitkomstinformatie kan worden gecombineerd met andere aspecten/waarden die voor de patiënt belangrijk zijn. Samen beslissen is een circulair en interdisciplinair proces. Het speelt zich af in meerdere gesprekken, door meerdere zorgprofessionals, en de patiënt kan op elk moment terugkomen op een eerder gemaakte keuze. Besluitvorming vraagt tijd: patiënten moeten de ruimte krijgen om (uitkomst)informatie te laten bezinken, thuis over opties na te denken en eventueel naasten te raadplegen voordat een keuze wordt gemaakt. De stappen in dit hoofdstuk zijn dan ook geen strikte volgorde. Het zijn handvatten die men flexibel kan inzetten, afhankelijk van de situatie en de context (acuut versus chronisch, kort versus lang zorgpad).

Uitkomstinformatie kan afkomstig zijn uit verschillende bronnen: dashboards met eigen uitkomsten (N=1) of gegevens van vergelijkbare patiënten ('patients like me'), keuzehulpen met populatiegegevens of trial data, voorspelmodellen, of rechtstreeks uit richtlijnen en literatuur.

Bij 'patients like me' is uitkomstinformatie vooral vooruitkijkend. Deze informatie laat zien wat je kunt verwachten van verschillende behandelingen en ondersteunt de afweging welke optie het beste past. N=1-uitkomstinformatie kijkt terug. Deze informatie kan helpen om een keuzemoment te signaleren, context te bieden bij het inschatten van verwachte effecten, of te ondersteunen bij de evaluatie of de huidige behandeling nog passend is. Tegelijkertijd dekt uitkomstinformatie niet alles wat voor een patiënt relevant kan zijn. De vragen in dit hoofdstuk helpen om uitkomstinformatie in te bedden in een breder gesprek over waarden en voorkeuren. Hierdoor ondersteunen de data het gezamenlijke besluitvormingsproces in plaats van het te domineren.

De vragen in dit hoofdstuk zijn georganiseerd langs de vier stappen van Samen beslissen:

Stap	Kern	Vragen
Stap 1: Keuze	Erkennen dat er een keuze is en dat perspectieven van patient en zorgprofessional er beide toe doen	3.1
Stap 2: Opties	Bespreken van (behandel)mogelijkheden en hoe deze scoren op relevante uitkomsten en aspecten	3.2 - 3.4
Stap 3: Voorkeur	Verkennen van voorkeuren, waarden en context van de patiënt, inclusief eventuele absolute grenzen	3.5
Stap 4: Beslissing	Gezamenlijk afwegen en tot een passende keuze komen	3.6 - 3.7

Methodologische basis

De vier stappen van Samen beslissen die het uitgangspunt vormen voor deze sectie sluiten aan bij het Three-Talk Model van Elwyn et al. (2012, 2017), waarin de fasen Team Talk, Option Talk en Decision Talk worden onderscheiden. Gekozen is voor vier stappen in plaats van drie, omdat dit model explicieter ruimte biedt voor het herkennen van het keuzemoment als aparte stap.

De aandacht voor het uitvragen van voorkeuren en gezamenlijk afwegen is gebaseerd op het werk van Stiggelbout et al. (2012, 2015). Zij benadrukken dat voorkeuren van patiënten breed variëren, vaak verschillen van wat zorgprofessionals verwachten, en niet voorspeld kunnen worden op basis van patiëntkenmerken. Een kernpunt in hun werk is dat voorkeuren vaak niet kant-en-klaar bestaan, maar worden gevormd tijdens het gesprek (preference construction). Dit betekent dat de zorgprofessional een actieve rol heeft in het ondersteunen van dit proces door open vragen te stellen en de patiënt te helpen verhelderen wat werkelijk belangrijk is. Stiggelbout et al. constateren dat in de praktijk deze afwegingsstap vaak wordt overgeslagen. zorgprofessionals presenteren informatie over opties, maar gaan vervolgens direct naar de beslissing zonder de voorkeuren van de patiënt expliciet te inventariseren en te bespreken.

Dit hoofdstuk presenteert een raamwerk dat zorgprofessionals en patiënten kan helpen om uitkomstinformatie op een gestructureerde manier te integreren in de gezamenlijke besluitvorming. De aanpak is geïnspireerd door het werk van Dolan (2010) over multicriteria clinical decision support: methoden die structuur bieden aan complexe klinische beslissingen met meerdere dimensies, waarbij patiënt-voorkeuren en -waarden centraal staan. Uit onderzoek blijkt dat vereenvoudigde methoden zoals rangorde-gebaseerde afwegingen in de meeste gevallen tot vergelijkbare conclusies leiden als complexere gewogen modellen. En ze passen beter bij de tijdsdruk en praktische beperkingen van de spreekkamer (Dolan et al., 2014). Dit raamwerk gebruikt de eenvoudige aanpak. Het werkt met duidelijke aspecten, kwalitatieve beoordelingen en een rangorde van opties, zonder ingewikkelde berekeningen of totaalscores.

Randvoorwaarden

Samen beslissen met uitkomstinformatie speelt zich af op drie niveaus. Het microniveau is de spreekkamer: het gesprek tussen patiënt en zorgprofessionals. Daar staan de vier stappen centraal die in de rest van dit hoofdstuk stap voor stap worden uitgewerkt. Maar of die stappen daadwerkelijk uitvoerbaar zijn, hangt sterk af van wat er op het meso- en macroniveau is georganiseerd: het ziekenhuis, het team, de beschikbare hulpmiddelen en de landelijke infrastructuur. Daarnaast verschilt de plaats die uitkomstinformatie kan innemen sterk per type zorgpad. Spoedzorg vraagt een andere invulling dan een chronisch of oncologisch traject.

Randvoorwaarden op meso- en macroniveau

Mesoniveau: ziekenhuisbeleid en teamorganisatie

1. Ziekenhuizen en vakgroepen hebben een verantwoordelijkheid om eenduidig beleid te voeren over welke keuzehulpen en informatiematerialen men inzet. Onderlinge afstemming is cruciaal: tegenstrijdige informatie van verschillende zorgprofessionals leidt tot verwarring bij patiënten. Het is belangrijk om aan te sluiten bij landelijk beleid waar dit er is.
2. Het helpt als op teamniveau is nagedacht over hoe de stappen van Samen Beslissen zijn verdeeld. Niet alle stappen hoeven door de medisch specialist te worden gezet. Andere zorgprofessionals kunnen een deel van de informatie voorbereiding en waardenverduidelijking op zich nemen.
3. Voor goede continuïteit is het van belang dat organisaties borgen dat de uitkomsten van Samen-beslisgesprekken worden vastgelegd in het elektronisch patiëntendossier (EPD). Zo zijn alle betrokken zorgprofessionals op de hoogte van eerder gemaakte afspraken en beslissingen.
4. Houd er rekening mee dat het duiden en inbedden van uitkomstinformatie in het gesprek specifieke vaardigheden vraagt die niet vanzelfsprekend onderdeel zijn van de basisopleiding. Opleiding en training op dit terrein is de verantwoordelijkheid van zorginstellingen en de zorgprofessional zelf.

Macroniveau: landelijke afspraken en infrastructuur

1. Op landelijk niveau is uniformiteit in uitkomstinformatie een belangrijke randvoorwaarde. Waar landelijke uitkomstensets beschikbaar zijn, verdient het aanbeveling deze te gebruiken. Waar deze (nog) ontbreken, is het aan de wetenschappelijke verenigingen om afspraken te maken over welke uitkomsten per aandoening relevant zijn en hoe deze worden gepresenteerd. Verschillende instellingen die tegenstrijdige informatie geven, ondermijnen het vertrouwen van patiënten.
2. Het is wenselijk dat er structurele financiering en een landelijk implementatieplan voor keuzehulpen komt, met een onderhoudscyclus gekoppeld aan richtlijnen, zodat de informatie die patiënten ontvangen consistent is met de meest actuele medische inzichten en alle relevante behandelopties omvat.

Samen beslissen en uitkomstinformatie per type zorgpad

Samen beslissen is in elk zorgpad van belang, maar de invulling ervan, en daarmee de rol en waarde van uitkomstinformatie, verschilt per type zorgpad. Onderstaand onderscheid in vier typen kan zorgprofessionals en teams helpen om de stappen uit dit hoofdstuk passend in te zetten.

1. **Spoedzorg.** Beslissingen moeten soms binnen minuten worden genomen, waardoor een uitgebreid Samen-beslisproces niet altijd mogelijk is. Uitkomstinformatie is hier vooral van waarde in de vorm van parate klinische kennis van de zorgprofessional (“wat is het verwachte beloop met of zonder behandeling?”). De nadruk ligt op het benoemen van de keuze, het kort schetsen van opties en gevolgen, en het respecteren van eerder vastgelegde wilsverklaringen of behandelgrenzen. Proactieve zorgplanning, beslissingen die vóór een acute situatie zijn besproken, vergroot de kwaliteit van Samen beslissen ook in spoedcontexten.
2. **Chronische ziekte.** Dit is het zorgpad waarvoor uitkomstinformatie het meest uitgewerkt beschikbaar is en het meest waardevol kan zijn. Er is tijd om keuzehulpen in te zetten, patiënten voor te bereiden, en het gesprek over meerdere contactmomenten te spreiden. Uitkomstinformatie speelt een rol bij meerdere typen beslissingen. De start van een behandeling, aanpassing bij verandering in ziektebeloop, en evaluatie van de huidige aanpak. Het circulaire karakter van Samen beslissen is hier het sterkst zichtbaar. Waarden en voorkeuren van patiënten kunnen over de tijd veranderen en verdienen regelmatige hertoetsing.
3. **Kort zorgpad (één of enkele contactmomenten).** De uitdaging is hier dat er weinig tijd en ruimte is voor voorbereiding en nabespreking. Een keuzehulp kan voorafgaand aan het eerste consult worden aangeboden, maar als er geen vervolgsconsult is, ontbreekt het moment om de afweging te verdiepen. In dit type zorgpad is het extra belangrijk om de patiënt actief te wijzen op beschikbare informatiebronnen en na het consult de mogelijkheid te bieden vragen te stellen (bijv. via e-consult of een terugbelmogelijkheid). Uitkomstinformatie is hier vaak generiek van aard en minder gepersonaliseerd.
4. **Lang zorgpad met meerdere behandelfasen (bijv. oncologie).** In dit type zorgpad spelen meerdere opeenvolgende beslismomenten, waarbij de context (ziekt stadium, behandelrespons, kwaliteit van leven) voortdurend verandert. Uitkomstinformatie heeft hier een bijzondere waarde: n=1-data (de eigen uitkomsten van de patiënt) kunnen signaleren wanneer een behandeling bijgesteld moet worden. Populatiegegevens helpen bij het inschatten van de te verwachten effecten van vervolgbehandelingen. De emotionele belasting is in dit type zorgpad vaak hoog. Dat vraagt extra aandacht voor de dosering en duiding van informatie en voor de vraag hoeveel regie de patiënt op een gegeven moment wil en kan dragen.

Grote en kleine beslissingen

Samen beslissen is niet voorbehouden aan grote behandelbeslissingen. Ook kleinere beslissingen in het zorgproces, bijvoorbeeld wel of geen aanvullend onderzoek, of zelfs praktische zaken zoals wil de patiënt liggend of zittend gewassen worden, zijn keuzemomenten. Het onderscheid tussen grote en kleine beslissingen is relevant voor de manier waarop de stappen uit dit hoofdstuk worden ingezet. Een grote behandelbeslissing vraagt doorgaans meer voorbereiding, meer tijd en mogelijk meerdere gesprekken. Een kleinere beslissing kan in een enkel kort gesprek worden doorlopen. De stappen blijven hetzelfde, maar de intensiteit verschilt.

Gezondheidsvaardigheden en inclusiviteit

Samen beslissen met uitkomstinformatie veronderstelt dat patiënten informatie kunnen begrijpen, verwerken en meewegen in een beslissing. Daarbij speelt de voorkennis van de patiënt een belangrijke rol. Om uitkomstinformatie goed te kunnen interpreteren, is het essentieel dat de patiënt begrijpt wat de aandoening inhoudt, welke effecten deze kan hebben op het functioneren en wat de gebruikte uitkomstmaten daarover zeggen. Voordat het proces van Samen Beslissen begint, moet daarom worden nagegaan of de patiënt over voldoende inzicht en een passend referentiekader beschikt. Als dit niet het geval is, dient eerst (meer) uitleg te worden gegeven.

Niet alle patiënten hebben dezelfde gezondheidsvaardigheden. In Nederland heeft naar schatting 25% van de bevolking beperkte gezondheidsvaardigheden (laaggeletterdheid, beperkte digitale vaardigheden of cognitieve beperkingen). Dit heeft directe consequenties voor hoe uitkomstinformatie wordt aangeboden en hoe Samen beslissen wordt vormgegeven. Inclusief Samen beslissen vraagt om bewuste keuzes in taal, vorm en proces.

Stap 1: Keuze

Patiënt en het zorgteam (h)erkennen samen dat er een keuze te maken is, en dat beide perspectieven belangrijk zijn voor het besluit.

3.1 (H)erkenning van het keuze moment

Is het voor de patiënt en zorgprofessional helder dat er samen een keuze gemaakt moet worden, en welke keuze dat is?

Aanbevelingen

1. Benoem in het gesprek expliciet dat er een keuze gemaakt moet/kan worden en dat de patiënt hierin een rol heeft. Een activerende zin zoals: “Er is een keuze te maken en die gaan we samen maken” kan hierbij helpen. Deze formulering maakt het gezamenlijke karakter van het beslisproces meteen concreet.
2. Geef aan dat er meerdere opties zijn en dat de beste keuze niet op voorhand vaststaat maar afhangt van wat voor de patiënt belangrijk is.
3. Wees er alert op dat het tonen van uitkomstinformatie de indruk kan wekken dat er weinig te kiezen valt; benadruk dat cijfers informatie geven over de keuze, maar de keuze niet bepalen.
4. Houd rekening met de context: in een acute situatie moet het keuzemoment soms direct worden benoemd en kan besluitvorming niet worden uitgesteld. In een chronische setting is er vaak meer tijd om het gesprek voor te bereiden en te spreiden over meerdere contactmomenten.

Toelichting

Samen beslissen begint met de gezamenlijke (h)erkenning dat er een keuze voorligt en dat zowel het perspectief van de zorgprofessional als van de patiënt ertoe doen. Dit lijkt vanzelfsprekend, maar in de praktijk is het keuzemoment niet altijd expliciet. Patiënten kunnen het gevoel hebben dat het behandeltraject al vaststaat, of niet beseffen dat er alternatieven zijn. Zorgprofessionals kunnen onbewust sturen naar een voorkeursoptie zonder de keuze als zodanig te benoemen. Bovendien is dit ‘herkennen’ van het keuzemoment niet uitsluitend de taak van de medisch specialist: andere zorgprofessionals vervullen hierin een even belangrijke rol, vaak in de aanloop náár het consultgesprek.

Wanneer uitkomstinformatie wordt gebruikt (of dit nu uit een dashboard, keuzehulp, voorspelmodel of andere bron komt) is het extra belangrijk om eerst het keuzemoment te markeren. Cijfers hebben een zekere autoriteit: ze kunnen de indruk wekken objectief en beslissend te zijn, waardoor de ruimte voor eigen afweging kleiner lijkt dan zij is. Onderzoek laat zien dat het presenteren van kwantitatieve informatie ertoe kan leiden dat zowel patiënten als zorgprofessionals minder ruimte ervaren voor persoonlijke voorkeuren en waarden. Benoem vooraf expliciet dat er een keuze is én dat die keuze afhangt van wat voor de patiënt belangrijk is. Hiermee voorkom je dat de data het gesprek domineren voordat het gesprek is begonnen.

Stap 2: Opties

De verschillende behandel- of zorgmogelijkheden worden besproken, inclusief de voor- en nadelen. Ook het afzien van behandeling kan een optie zijn.

3.2 Beschikbare behandelopties

Welke behandelopties zijn beschikbaar en reëel (inclusief niet curatief behandelen)?

Aanbevelingen

1. Benoem alle medisch verantwoorde opties expliciet, inclusief afwachtend beleid, niet curatief behandelen of geen verdere diagnostiek. Overweeg om 'niet curatief behandelen' als eerste optie te introduceren, samen met het natuurlijke beloop van de aandoening. Dit geeft patiënten een helder referentiepunt voor de overige opties.
2. Wees transparant als bepaalde opties lokaal niet beschikbaar zijn maar elders wel.
3. Sluit geen opties vooraf uit op basis van aannames over patiëntvoorkeuren (je kunt als zorgprofessional niet vooraf weten wat voor een patiënt wel of niet belangrijk is). Bij een groot aantal theoretisch mogelijke opties is een voorselectie soms legitiem. Maak deze op basis van de medische en persoonlijke situatie van de patiënt. Leg aan de patiënt uit waarom bepaalde opties niet besproken worden.
4. Betrek de uitkomsten van een multidisciplinair overleg (MDO) als input en leg dit voor aan de patiënt. Het MDO-voorstel hoeft niet automatisch leidend zijn.

Toelichting

Voordat uitkomst informatie kan worden besproken, moet helder zijn welke opties daadwerkelijk op tafel liggen. Soms worden opties bewust weggelaten om ze niet volgens de richtlijn zijn, of omdat ze in de eigen instelling niet beschikbaar zijn. Ook conservatief beleid, afwachtend beleid, of afzien van curatieve behandeling verdient expliciet een plek als realistische optie.

Het is belangrijk om de beschikbare opties te inventariseren onafhankelijk van de vraag of daar uitkomst informatie over beschikbaar is. Dashboards, keuzehulp en voorspelmodellen bevatten vaak een vaste set opties die niet voor elke patiënt allemaal reëel zijn. Of ze missen juist opties die voor de betreffende patiënt wél relevant zijn. De zorgprofessional bepaalt eerst welke behandelopties voor deze specifieke patiënt daadwerkelijk beschikbaar zijn. Zo voorkom je dat de keuze wordt gestuurd door de opties die toevallig in dashboards, keuzehulp of voorspelmodellen staan.

3.3 Relevante aspecten voor de patiënt

Welke aspecten zijn voor de patiënt relevant bij deze keuze?

Aanbevelingen

1. Inventariseer tijdens het gesprek welke aspecten voor de patiënt meewegen bij deze keuze. Stel open vragen zoals “Wat is voor u belangrijk bij deze keuze?” en help de patiënt te verhelderen wat werkelijk belangrijk is voor de patiënt door door te vragen.
2. Breng zowel gezondheidsuitkomsten (zoals pijn, functioneren, kwaliteit van leven) als andere aspecten (zoals behandelbelasting, reistijd, impact op werk of gezin) expliciet in kaart.
3. Koppel waar relevant de aspecten die de patiënt noemt aan meetbare uitkomsten. Bijvoorbeeld: als een patiënt zegt “ik wil kunnen blijven wandelen”, kan de zorgprofessional dit koppelen aan uitkomsten over functioneren of mobiliteit.
4. Nodig patiënten actief uit om zich voor te bereiden op het gesprek, eventueel samen met een naaste. Wijs patiënten hiervoor actief op beschikbare hulpmiddelen (keuzehulp, keuzekaarten, informatie filmpjes). Informeer ook over wat uitkomst informatie is en wat het niet zegt, zodat patiënten cijfers in de juiste context kunnen plaatsen.

5. Bepaal samen met de patiënt welke aspecten belangrijk zijn voor de patiënt. Dit kan mondeling plaatsvinden in een gesprek, maar er kan ook gebruik gemaakt worden van een afwegingsoverzicht: een tabel met de behandelopties als kolommen en de aspecten als rijen. Deze tabel kan dienen als hulpmiddel voor een gesprek, maar is geen vereiste.
6. Erken dat aspecten waarvoor geen kwantitatieve data beschikbaar zijn even legitiem zijn als uitkomsten met cijfers.

Toelichting

Bij elke behandelkeuze spelen verschillende soorten overwegingen een rol. Enerzijds zijn er gezondheidsuitkomsten: meetbare resultaten van een behandeling zoals pijn, functioneren, kwaliteit van leven of overleving. Anderzijds zijn er andere aspecten die voor patiënten minstens zo belangrijk kunnen zijn. Zoals de belasting van de behandeling, frequentie van ziekenhuisbezoek, impact op werk of gezin, of de wens om dicht bij naasten te blijven.

Systematische reviews naar patiëntvoorkeuren bij behandelkeuzes laten zien dat gestructureerde tools zoals keuzehulpen aspecten doorgaans indelen in uitkomsten, procesaspecten en kosten. Welke categorie het belangrijkste gevonden wordt varieert echter sterk per context en per studie. Bovendien blijkt dat zorgprofessionals en patiënten soms verschillende prioriteiten toekennen aan dezelfde aspecten. Kwalitatief onderzoek laat daarnaast zien dat patiënten ook factoren meewegen die niet in vooraf gedefinieerde lijsten passen. Zoals persoonlijke levensdoelen en filosofie, sociaal-culturele achtergrond, situationele factoren, en relationele waarden zoals vertrouwen in de zorgprofessional. Dit sluit aan bij het inzicht van Stiggelbout en collega's dat voorkeuren niet kant-en-klaar bestaan maar tijdens het gesprek worden gevormd. Dat onderstreept het belang van een open dialoog waarin de zorgverlener actief inventariseert wat voor deze specifieke patiënt en in deze specifieke situatie belangrijk is.

Het inventariseren van voor de patiënt relevante aspecten vraagt om een actieve, ondersteunende houding van de zorgprofessional. Open vragen gericht op wat voor de patiënt belangrijke afwegingen zijn bij deze keuze. De vragen die de zorgprofessional stelt en de informatie die zorgprofessional en patiënt bespreken dragen bij aan het vormen van voorkeuren van de patiënt.

Het is de taak van de zorgprofessional om de aspecten die de patiënt noemt waar nodig en mogelijk te koppelen aan uitkomsten waarvoor informatie beschikbaar is. Een patiënt zal zelden zeggen "fysiek functioneren is belangrijk voor mij", maar wel "ik wil met mijn kleinkinderen kunnen spelen". De zorgprofessional herkent dit als een vraag over functioneren en kan dit koppelen aan beschikbare uitkomst-informatie.

Aanvullend kan overwogen worden om patiënten voorafgaand aan het consult te vragen na te denken over wat zij belangrijk vinden. Hiervoor kunnen beschikbare hulpmiddelen worden ingezet zoals keuzehulpen, vragenlijsten of informatie-video's. Voorbereiding geeft patiënten tijd om in een rustige omgeving na te

denken, wat het gesprek kan verdiepen. Voorbereiding vervangt echter niet het gesprek. De uiteindelijke verheldering en nuancering van voorkeuren vindt plaats in de interactie met de zorgprofessional.

Het opschrijven van de belangrijke aspecten voor de patiënt dient als geheugensteun. Het kan ervoor zorgen dat alle aspecten die de patiënt belangrijk vindt daadwerkelijk worden meegenomen in de afweging. Een tabel met opties als kolommen en aspecten als rijen kan hierbij dienen als hulpmiddel (Figuur 1).

Aspect	Belang patiënt	Behandeling 1	Behandeling 2
Effectiviteit	★ ★ ★	++	+-
Bijwerkingen			
Impact op sociale activiteiten			
....			

Figuur 1: Voorbeeld van een afwegingsoverzicht met de behandelopties in de kolommen (dit kunnen er meer zijn dan 2) en de belangrijke aspecten in de rijen.

3.4 Vergelijking van opties op relevante aspecten

Hoe verhouden de verschillende behandelopties zich tot elkaar op de relevante uitkomsten en aspecten?

Aanbevelingen

1. Bespreek samen welke opties op welke aspecten beter of slechter scoren. Dit kan mondeling, of ondersteund door het afwegingsoverzicht. Als je het overzicht gebruikt: begin met de aspecten waarvoor geen uitkomst informatie nodig is, zoals behandelbelasting, frequentie van ziekenhuisbezoek en impact op dagelijks leven. De zorgprofessional kan deze vaak direct scoren op basis van parate kennis.
2. Raadpleeg de beschikbare uitkomst informatie voor de gezondheidsuitkomsten die de patiënt belangrijk vindt. Bespreek samen wat de cijfers betekenen in relatie tot wat de patiënt belangrijk vindt.

3. Bij gebruik van het overzicht: scoor samen met de patiënt elke optie op elk aspect (bij gebruik van het overzicht). Gebruik een eenvoudige notatie (bijvoorbeeld: ++, +, 0, -, --) of steekwoorden (zoals 'lang', 'weinig', 'ernstig') zodat in één oogopslag zichtbaar is hoe opties zich tot elkaar verhouden.
4. Vertaal kwantitatieve uitkomstinformatie samen met de patiënt naar de gekozen notatie. Bespreek wat de cijfers voor hem of haar betekenen: een voorspelde pijnreductie van 40% kan voor de ene patiënt een ++ zijn, voor een andere patiënt slechts een +.
5. Controleer of de uitkomstinformatie alle geïnventariseerde opties dekt. Als bepaalde opties ontbreken, schat de scores in op basis van klinische kennis.
6. Vermijd schijnprecisie: een kwalitatieve inschatting (++/+/0/-/--) is transparanter dan een geforceerd cijfer.

Toelichting

Het invullen van het afwegingsoverzicht is één van de manieren waarop de vergelijking van opties vorm kan krijgen. Wanneer men kiest voor dit hulpmiddel, verloopt het invullen doorgaans in twee fasen. In de eerste fase worden de aspecten gescoord waarvoor geen uitkomstinformatie nodig is. Dit zijn vaak de praktische en persoonlijke aspecten: behandelbelasting, reistijd, impact op werk en/of gezin en hersteltijd. De zorgprofessional kan deze aspecten meestal direct scoren op basis van parate kennis over de behandelingen. Soms is input van de patiënt nodig, bijvoorbeeld bij de vraag hoe zwaar een bepaalde belasting weegt in de specifieke situatie van de patiënt.

In de tweede fase raadpleegt de zorgprofessional samen met de patiënt de uitkomstinformatie voor de gezondheidsuitkomsten die voor de patiënt belangrijk zijn. Deze informatie komt bijvoorbeeld uit een dashboard, keuzehulp, voorspel model of een andere bron.

Het scoren van uitkomstinformatie is geen mechanische vertaling van cijfers naar plusjes en minnen. Wat een bepaalde uitkomst betekent, hangt af van de patiënt. Een voorspelde pijnreductie van 40% kan voor een patiënt met ernstige pijn een grote verbetering zijn (++), terwijl dezelfde reductie voor een patiënt met milde pijn minder relevant is (+). Door dit samen te bespreken, worden de scores betekenisvol voor de specifieke situatie.

Het voordeel van deze aanpak is dat alle aspecten - zowel met als zonder kwantitatieve onderbouwing - op een vergelijkbare manier zichtbaar worden. Dit voorkomt dat kwantitatieve uitkomsten automatisch zwaarder wegen, simpelweg omdat er cijfers bij staan. De notatie kan worden aangepast aan de situatie: bij een beperkt aantal aspecten kunnen steekwoorden ('lange wachttijd', 'veel pijnreductie', 'dagelijks medicatie') soms helderder zijn dan abstracte symbolen. Tegelijkertijd blijft het een hulpmiddel voor het gesprek, geen algoritme dat de beslissing dicteert.

Het ingevulde afwegingsoverzicht kan de patiënt meenemen om thuis verder te overdenken, te bespreken met naasten, of als basis voor een vervolgesprek. Dit is vooral waardevol bij complexe beslissingen waarbij bedenktijd nodig is.

Stap 3: Voorkeur

De patiënt vertelt over persoonlijke waarden, wensen en omstandigheden. De zorgprofessional koppelt deze aan de beschikbare medische mogelijkheden.

Het doseren van uitkomstinformatie is hier cruciaal: te veel informatie tegelijk leidt niet tot betere beslissingen. Uitkomstcijfers mogen nooit zonder duiding worden aangeboden, een zorgprofessional helpt de patiënt de betekenis van cijfers te begrijpen in de eigen situatie. Sommige patiënten willen zelf de beslissing nemen, anderen geven de voorkeur aan een advies van de zorgprofessional. Beide zijn legitiem; de zorgprofessional stemt zijn of haar rol hierop af.

3.5 Absolute grenzen en dealbreakers

Zijn er aspecten waarop een optie zo slecht scoort dat deze voor de patiënt niet acceptabel is, ongeacht de scores op andere aspecten?

Aanbevelingen

1. Vraag expliciet naar absolute grenzen voordat de afweging begint: “Zijn er uitkomsten of situaties die voor u echt niet acceptabel zijn?” Soms is het moeilijk om te zeggen wat iemand per se niet wil, maar de meeste mensen kunnen nog wel aangeven wat ze minimaal nog willen kunnen. De vraag: “Wat zou u niet willen missen, wat wilt u absoluut nog kunnen blijven doen?” kan hierbij helpend zijn.
2. Vraag gericht door als een optie op een cruciaal aspect erg slecht scoort: “Is dit voor u een harde grens, of iets wat u nog zou kunnen accepteren?”
3. Maak onderscheid tussen sterke voorkeuren (die kunnen worden afgewogen tegen andere aspecten) en absolute grenzen (die een optie uitsluiten, ongeacht hoe goed deze op andere aspecten scoort).
4. Sluit opties met dealbreakers expliciet uit voordat de verdere afweging plaatsvindt.
5. Let op dat kwantitatieve uitkomsten niet automatisch zwaarder wegen alleen omdat er cijfers bij staan; een dealbreaker op een niet-gemeten aspect is even valide.

Toelichting

Naast voorkeuren die tegen elkaar kunnen worden afgewogen, kunnen er ook absolute grenzen zijn: uitkomsten of behandelkenmerken die voor een patiënt zo onacceptabel zijn dat ze een optie uitsluiten. Dit kunnen ernstige bijwerkingen zijn, maar ook niet-medische zaken zoals langdurige opname, frequente ziekenhuisbezoeken, of specifieke beperkingen in dagelijks functioneren.

Het afwegingsoverzicht maakt dergelijke dealbreakers zichtbaar: als een optie op een aspect '--' scoort en de patiënt geeft aan dat dit niet acceptabel is, dan valt deze optie af. Dit vereenvoudigt de daaropvolgende afweging, omdat zorgprofessional en patiënt alleen de resterende opties hoeven te vergelijken.

Het is belangrijk om dealbreakers te onderscheiden van sterke voorkeuren. Een sterke voorkeur kan worden afgewogen: een optie die slecht scoort op één aspect kan nog steeds de beste zijn als deze uitblinkt op belangrijkere aspecten. Een dealbreaker daarentegen is absoluut: de optie valt af, ongeacht de scores elders. Door dit onderscheid expliciet te maken, voorkom je dat je opties onterecht uitsluit of juist onterecht in de afweging houdt.

Stap 4: Beslissing

Patiënt en zorgprofessional nemen gezamenlijk een besluit door medische kennis en persoonlijke voorkeuren ten opzichte van de opties samen af te wegen.

3.6 Gezamenlijke afweging

Hoe komen patiënt en zorgprofessional samen tot een afgewogen keuze die past bij wat voor de patiënt belangrijk is?

Aanbevelingen

1. Ondersteun de patiënt actief bij het verhelderen van voorkeuren. Voorkeuren worden vaak pas duidelijk door het gesprek zelf.
2. Vraag de patiënt welke aspecten het zwaarst wegen: “Wat is voor u het allerbelangrijkst bij deze keuze? En daarna?” Dit kan men mondeling bespreken of vastleggen in het overzicht als dat wordt gebruikt.
3. Als het afwegingsoverzicht wordt gebruikt: leg de rangorde van aspecten daarin vast in een aparte kolom. Dit maakt zichtbaar welke optie het beste aansluit bij wat voor de patiënt het zwaarst weegt.

4. Breng de aspecten en opties bij elkaar, in gesprek of via het overzicht, en zoek samen naar welke optie het beste aansluit bij wat voor de patiënt het zwaarst weegt.
5. Identificeer waar spanning zit: scoort één optie het beste op het belangrijkste aspect, maar slechter op andere hoog-geprioriteerde aspecten? Bespreek deze afweging expliciet.
6. Als er geen duidelijke winnaar is, benoem dit en bespreek hoe verder te komen: bedenktijd nemen, aanvullende informatie zoeken, of een gesprek met naasten.
7. Het afwegingsoverzicht is één van de vormen die kan worden gebruikt om de afweging te structureren, het is geen beslialgoritme en geen vereiste. De afweging kan ook volledig mondeling plaatsvinden. Kies de vorm die past bij de patiënt, de situatie en de beschikbare tijd. De waarde zit altijd in het gesprek zelf, niet in het invullen van een overzicht.
8. Zorg dat de verslaglegging van de afweging en het besluit toegankelijk is voor alle betrokken zorgprofessionals en, waar van toepassing, ook voor de patiënt zelf. Continuïteit in het dossier maakt het mogelijk om bij een volgend contactmoment voort te bouwen op eerdere gesprekken.

Toelichting

Gezamenlijke afweging is het kernproces waarin voorkeuren worden verhelderd en afgewogen. Onderzoek laat zien dat deze stap in de praktijk vaak wordt overgeslagen: zorgprofessionals presenteren informatie over opties en gaan om verschillende redenen vervolgens direct naar de beslissing, zonder de voorkeuren van de patiënt expliciet te bespreken. Dit leidt ertoe dat beslissingen niet altijd aansluiten bij wat voor de patiënt werkelijk belangrijk is.

en belangrijk inzicht is dat voorkeuren vaak niet vooraf bestaan maar worden gevormd tijdens het gesprek. De zorgprofessional heeft daarom een actieve rol. Niet passief afwachten wat de patiënt wil, maar helpen verhelderen door open vragen te stellen, door te vragen en ruimte te geven voor twijfel en nuance. Dit proces van voorkeursvorming (preference construction) vraagt tijd en aandacht.

Het afwegingsoverzicht kan als hulpmiddel worden ingezet om de afweging te structureren en zichtbaar te maken, maar het is niet de enige manier. De kern van stap 4 is het gesprek zelf. Zorgprofessional en patiënt redeneren samen hardop, brengen waarden en opties bij elkaar, en zoeken naar wat voor deze patiënt, in deze situatie, het beste past. Dat kan mondeling gebeuren, eventueel ondersteund door een eenvoudig overzicht op papier of scherm. Wanneer een afwegingsoverzicht wordt gebruikt, helpt het om te beginnen bij het aspect dat voor de patiënt het zwaarst weegt, en van daaruit de overige aspecten te betrekken.

Deze aanpak maakt spanning zichtbaar: als de optie die het beste scoort op het belangrijkste aspect slechter scoort op het tweede of derde aspect, is dat precies waar de afweging zit. Dit zijn de momenten die vragen om verdere afweging: wat weegt zwaarder, en waarom? Of er nu wel of geen overzicht op tafel ligt, dit soort spanning is alleen te doordenken in een dialoog. Het hardop redeneren mét de patiënt maakt impliciete aannames zichtbaar en helpt voorkeuren te verhelderen. Het systematisch doorlopen van opties, aspecten en prioriteiten dient niet om een berekend 'beste antwoord' te genereren, maar om het gesprek te structureren en blinde vlekken te voorkomen. De waarde zit in de deliberatie zelf: door hardop te redeneren worden impliciete aannames zichtbaar en kan de patiënt zijn of haar eigen voorkeuren scherper krijgen.

Deze aanpak sluit aan bij wat Dolan (2010) 'multi-criteria clinical decision support' noemt: methoden die structuur bieden aan complexe klinische beslissingen met meerdere dimensies, waarbij patiëntvoorkeuren en -waarden centraal staan. Onderzoek wijst uit dat vereenvoudigde methoden zoals rangorde-gebaseerde afwegingen in de meeste gevallen tot vergelijkbare conclusies leiden als complexere gewogen modellen (Dolan et al., 2014). Dit terwijl ze beter passen bij de tijdsdruk en praktische beperkingen van de spreekkamer. Door de afweging aan het gesprek over te laten in plaats van aan een algoritme, wordt de patiënt ondersteund in het vormen van een geïnformeerde voorkeur die past bij zijn of haar persoonlijke situatie.

3.7 Congruentie met waarden en voorkeuren

Past de uiteindelijke keuze bij de waarden en voorkeuren van de patiënt?

Aanbevelingen

1. Check of de keuze aansluit bij de voorkeuren van de patiënt door bijvoorbeeld de patiënt in eigen woorden te laten uitleggen waarom deze keuze bij hem of haar past.
2. Wees alert op discrepanties tussen eerdergenoemde voorkeuren en de uiteindelijke keuze.
3. Als er discrepanties zijn, bespreek deze openlijk.

4. Bied ruimte om terug te komen op de keuze. Een eenmaal genomen besluit kan later worden herzien of bijgesteld op basis van hoe de behandeling in de praktijk uitpakt, of als omstandigheden en waarden van de patiënt veranderen. In een chronisch of lang zorgpad verdient dit een vast moment in het zorgpad. In kortere trajecten kan dit via een e-consult of terugbelmogelijkheid worden gefaciliteerd. Benoem deze mogelijkheid expliciet bij grote beslissingen.

Toelichting

Als afsluitende check is het waardevol om expliciet te toetsen of de gemaakte keuze congruent is met wat de patiënt eerder heeft aangegeven. Dit helpt voorkomen dat patiënten in sommige gevallen een suggestie van de zorgprofessional volgen die mogelijk niet het beste aansluit bij hun eigen voorkeuren en situatie. Het voorkomt ook dat de patiënt zich laat leiden door de meest prominente informatie in het dashboard, in plaats van door de eigen voorkeuren. Uit onderzoek blijkt dat patiënten die meer autonomie ervaren in de besluitvorming een grotere kans hebben op therapietrouw en sneller herstel. Dit onderstreept dat een goede congruentiecheck niet alleen ethisch gewenst is, maar ook klinisch relevant.

Referenties

- Barr, P. J., Thompson, R., Walsh, T., Grande, S. W., Ozanne, E. M., & Elwyn, G. (2014). The psychometric properties of CollaboRATE: a fast and frugal patient-reported measure of the shared decision-making process. *Journal of Medical Internet Research*, 16(1), e2.
- Bastemeijer, C. M., Voogt, L., van Ewijk, J. P., & Hazelzet, J. A. (2017). What do patient values and preferences mean? A taxonomy based on a systematic review of qualitative papers. *Patient Education and Counseling*, 100(5), 871-881.
- Bien, D. R., Danner, M., Vennedey, V., Civello, D., Evers, S. M., & Hiligsmann, M. (2017). Patients' preferences for outcome, process and cost attributes in cancer treatment: A systematic review of discrete choice experiments. *The Patient*, 10(5), 553-565.
- Dolan, J. G. (2010). Multi-criteria clinical decision support: A primer on the use of multiple-criteria decision-making methods to promote evidence-based, patient-centered healthcare. *Patient*, 3(4), 229-248.
- Dolan, J. G., Boohaker, E., Allison, J., & Imperiale, T. F. (2014). Can streamlined multi-criteria decision analysis be used to implement shared decision making for colorectal cancer screening? *MDM Policy & Practice*, 1(1), 1-11.
- Elwyn, G., Edwards, A., Wensing, M., Hood, K., Atwell, C., & Grol, R. (2003). Shared decision making: developing the OPTION scale for measuring patient involvement. *Quality and Safety in Health Care*, 12(2), 93-99.
- Elwyn, G., Frosch, D., Thomson, R., Joseph-Williams, N., Lloyd, A., Kinnersley, P., ... & Barry, M. (2012). Shared decision making: a model for clinical practice. *Journal of General Internal Medicine*, 27(10), 1361-1367.
- Elwyn, G., Lloyd, A., Joseph-Williams, N., Cording, E., Thomson, R., Durand, M. A., & Edwards, A. (2013). Option Grids: shared decision making made easier. *Patient Education and Counseling*, 90(2), 207-212.
- Elwyn, G., Lloyd, A., May, C., van der Weijden, T., Stiggelbout, A., Edwards, A., ... & Epstein, R. (2014). Collaborative deliberation: A model for patient care. *Patient Education and Counseling*, 97(2), 158-164.
- Elwyn, G., Durand, M. A., Song, J., Aarts, J., Barr, P. J., Berger, Z., ... & Van der Weijden, T. (2017). A three-talk model for shared decision making: multistage consultation process. *BMJ*, 359, j4891.
- Engels, N., van der Nat, P. B., Ankersmid, J. W., Prick, J. C. M., Parent, E., The, R., ... & van den Dorpel, M. A. (2022). Development of an online patient decision aid for kidney failure treatment modality decisions. *BMC Nephrology*, 23, 245.
- Essén, A., & Oborn, E. (2017). The performativity of numbers in illness management: The case of Swedish Rheumatology. *Social Science & Medicine*, 184, 134-143.
- Harrison, M., Milbers, K., Hudson, M., & Bansback, N. (2017). Do patients and health care providers have discordant preferences about which aspects of treatments matter most? Evidence from a systematic review of discrete choice experiments. *BMJ Open*, 7(5), e014719.
- ICHOM - International Consortium for Health Outcomes Measurement. <https://www.ichom.org/standard-sets/>
- Lee, Y. K., Low, W. Y., & Ng, C. J. (2013). Exploring patient values in medical decision making: A qualitative study. *PLoS ONE*, 8(11), e80051.
- Mullen, P. M. (2004). Quantifying priorities in healthcare: transparency or illusion? *Health Services Management Research*, 17(1), 47-58.
- Stacey, D., Carley, M., Gunderson, J., Hsieh, S. C., Kelly, S. E., Lewis, K. B., ... & Wells, G. (2025). The effect of patient decision aid attributes on patient outcomes: A network meta-analysis of a systematic review. *Medical Decision Making*, 45(3), 272-285.
- Stiggelbout, A. M., Van der Weijden, T., De Wit, M. P., Frosch, D., Légaré, F., Montori, V. M., ... & Elwyn, G. (2012). Shared decision making: really putting patients at the centre of healthcare. *BMJ*, 344, e256.
- Stiggelbout, A. M., Pieterse, A. H., & De Haes, J. C. (2015). Shared decision making: concepts, evidence, and practice. *Patient Education and Counseling*, 98(10), 1172-1179.
- Stone, D. (2020). *Counting: How We Use Numbers to Decide What Matters*. Liveright Publishing.
- Tünneßen, M., Hiligsmann, M., Stock, S., & Vennedey, V. (2020). Patients' preferences for the treatment of anxiety and depressive disorders: A systematic review of discrete choice experiments. *Journal of Medical Economics*, 23(6), 546-556.

4. Hoe verzamel en duid je uitkomstinformatie: N=1

Inleiding

Dit hoofdstuk beschrijft methodologische aandachtspunten voor het verantwoord gebruik van N=1-uitkomstinformatie. Bij het interpreteren van individuele uitkomsten spelen drie samenhangende aspecten een rol: (1) de interpretatie van scores en veranderingen, gegeven meetfout en onzekerheid; (2) de voorwaarden waaronder data tot stand komen, zoals meetmoment en consistentie; en (3) de manier waarop uitkomsten worden gepresenteerd.

De aanbevelingen richten zich primair op partijen die uitkomstinformatiertools ontwikkelen of beschikbaar stellen. Voor zorgprofessionals vormen zij achtergrondinformatie bij het beoordelen en gebruiken van dergelijke tools in de spreekkamer.

Praktische randvoorwaarden voor N=1-terugkoppeling in de spreekkamer

Een belangrijke praktische randvoorwaarde voor het gebruik van N=1-uitkomstinformatie in de spreekkamer is de beschikbaarheid van actuele data. In de huidige praktijk worden gegevens uit het elektronisch patiëntendossier (EPD) vaak met vertraging doorgeleverd aan kwaliteitsregistraties. Hierdoor zijn registratiedata veelal niet geschikt voor directe, individuele terugkoppeling tijdens het consult. Voor N=1-toepassingen is daarom real-time of bijna real-time dataverzameling noodzakelijk, bij voorkeur direct vanuit het EPD of via patiëntgerapporteerde metingen. Dit vraagt om bewuste keuzes en inrichting binnen zorginstellingen zelf, waarin processen, systemen en verantwoordelijkheden zodanig zijn georganiseerd dat actuele uitkomstinformatie daadwerkelijk beschikbaar is op het moment van het gesprek in de spreekkamer.

4.1 Relevantie van de uitkomstinformatie

Is de uitkomstinformatie relevant voor de patiënt én de zorgsituatie waarin deze wordt gebruikt?

Aanbevelingen

1. Zet dashboards of tools met uitkomstinformatie alleen in wanneer vooraf helder is welke vraag of behoefte in het spreekkamergesprek hiermee wordt ondersteund. Uitkomstinformatie is een middel om een bestaand gespreksdoel te faciliteren, niet een doel op zich.
2. Maak waar mogelijk gebruik van uitkomsten die patiënten zelf prioriteren. Het kan waardevol zijn om de prioriteit die patiënten aan uitkomsten toekennen te toetsen bij de eigen patiëntenpopulatie of aan de hand van kwalitatieve literatuur. Dit omdat bekend is dat perspectieven van patiënten en zorgprofessionals hierover kunnen verschillen.
3. Geef patiënten bij potentieel confronterende uitkomsten, zoals ernstige complicaties, de keuze of zij deze informatie willen zien. Niet elke patiënt wil of kan dergelijke informatie verwerken. Elke patiënt heeft wel het recht om te kiezen of deze wordt getoond.
4. Worden er klinische uitkomsten gebruikt die patiënten niet zelf voelen of merken en die maar indirect iets zeggen over wat voor patiënten belangrijk is? Maak dan expliciet duidelijk wat deze uitkomst betekent en waarom zij in deze situatie relevant is.

Toelichting

N=1-uitkomstinformatie is alleen bruikbaar wanneer de getoonde uitkomsten inhoudelijk relevant zijn voor zowel de patiënt als voor het zorgdoel of beslis-moment waarvoor zij worden ingezet. In de spreekkamer spelen uitkomsten van de patiënt vaak een bepalende rol bij het al dan niet voortzetten, aanpassen of intensiveren van behandeling. Dit vraagt om een bewuste en doelgerichte selectie van uitkomsten.

Enerzijds zijn er ziekte- en behandel specifieke uitkomsten die het klinisch beloop van een aandoening of het effect van zorg beschrijven, zoals labwaarden, functiemetingen of ziektespecifieke scores. Deze uitkomsten kunnen relevant zijn, maar de relatie met de gevolgen die voor de patiënt daadwerkelijk belangrijk zijn, is niet altijd direct duidelijk. Daarom is het nodig om expliciet te maken wat deze cijfers in de praktijk betekenen voor het dagelijks leven, de gezondheid en de voorkeuren van de patiënt.

Anderzijds zijn er patiëntrelevante uitkomsten, die beschrijven wat de impact van ziekte en zorg is op het dagelijks leven, zoals klachten, functioneren, kwaliteit van leven en - in sommige contexten - overleving. Deze uitkomsten worden, anders dan overleving, vaak gemeten met patiëntgerapporteerde uitkomsten (PROMs). Welke domeinen hierbij het meest relevant zijn, verschilt per aandoening en per patiënt.

Voor zinvol gebruik in de spreekkamer is het essentieel dat uitkomsten óf direct informatief zijn voor het behandelbeleid, óf een andere, duidelijk gedefinieerde functie vervullen binnen het zorgproces, zoals het signaleren van problemen of het monitoren van beloop. Wanneer dit niet het geval is, draagt het gebruik van uitkomstinformatie weinig bij aan het gesprek. Het kan zelfs afleiden van wat voor de patiënt en het beleid werkelijk van belang is.

4.2 Meetmomenten en timing

Op welke momenten in het zorgtraject wordt gemeten, en is er voldoende tijd tussen metingen om verandering te kunnen detecteren?

Aanbevelingen

1. Definieer bij voorkeur een vast startmoment voor metingen, zodat het beloop interpreteerbaar is en vergelijking over tijd mogelijk wordt. Een gemiste eerste meting betekent niet dat vervolgmetingen zinloos zijn. Juist bij chronische aandoeningen of langdurige behandelingen kan beloopinformatie zonder baselinemeting nog steeds waardevolle inzichten geven.
2. Plan meetmomenten rondom klinisch relevante momenten in het zorgtraject, zoals voor en na een ingreep, bij evaluatie van behandeling, of bij verwachte transitie momenten. Meet/registreer niet meer dan nodig is om de zorg te kunnen evalueren om registratielast niet onnodig te verhogen.
3. Zorg voor voldoende tijd tussen meetmomenten om daadwerkelijke verandering te kunnen detecteren. Raadpleeg indien nodig klinische trials of observationele studies voor het typische beloop van de uitkomst in vergelijkbare populaties.
4. Wanneer meetmomenten ontbreken, overweeg interpolatie om het beloop te schatten, maar wees transparant over de onzekerheid die dit geeft.
5. Bepaal bij het ontwerpen van het meetschema vooraf tot wanneer meten zinvol is. Wanneer na een bepaald moment in het zorgtraject geen relevante nieuwe informatie meer te verwachten is, hoeven vanaf dat punt geen data meer verzameld te worden.

Toelichting

De waarde van N=1-uitkomstinformatie hangt sterk af van wanneer wordt gemeten. Verandering wordt doorgaans geïnterpreteerd ten opzichte van een baseline-meting; zonder een consistent startmoment is het beloop daarom niet interpreteerbaar. Een vast startmoment, bijvoorbeeld bij diagnose, voor aanvang van behandeling, of bij instroom in een zorgprogramma, maakt vergelijking over tijd mogelijk.

Meetmomenten moeten aansluiten bij het zorgtraject. Meten vlak voor en na een ingreep geeft inzicht in het directe effect; meten bij geplande evaluatiemomenten ondersteunt beslissingen over voortzetten of aanpassen van beleid. Tegelijkertijd moet er voldoende tijd tussen metingen zitten om daadwerkelijke verandering te kunnen waarnemen. Als de verwachte verandering pas na maanden optreedt, is wekelijks meten van weinig toegevoegde waarde. Het kan mogelijk belastend zijn voor de patiënt of kostbaar voor de organisatie.

Voor het bepalen van een passend meetschema kan informatie uit klinische trials of observationele studies richting geven. Deze bronnen laten vaak zien wat het typische beloop is van een uitkomst na een bepaalde interventie of in een bepaald ziektestadium.

Wanneer meetmomenten ontbreken, bijvoorbeeld door non-respons of logistieke redenen, kan men interpolatie overwegen om het beloop te schatten. Dit introduceert echter onzekerheid die expliciet moet worden benoemd. Tot slot verdient het aanbeveling om bij het definiëren van het meetschema vooraf te bepalen tot wanneer meten zinvol is. Wanneer na een bepaald moment in het zorgtraject geen relevante nieuwe informatie meer te verwachten is, kan men besluiten om vanaf dat punt geen data meer te verzamelen.

4.3 Validiteit van individuele metingen

In hoeverre geeft de gemeten waarde een valide beeld van wat men beoogt te meten? Welke vertekeningen kunnen hierbij optreden?

Aanbevelingen

1. Houd er rekening mee dat veel vormen van vertekening minder problematisch zijn bij het interpreteren van beloop, mits de vertekening consistent is over meetmomenten. Bij vergelijking met afkapwaarden, referentiegroepen of cross-sectionele interpretatie is vertekening echter wel degelijk relevant.
2. Voor patiëntgerapporteerde uitkomsten: kies een instrument met een geschikte recallperiode (bij voorkeur kort of geen recallperiode). Meet op momenten die aansluiten bij deze periode, zodat patiënten niet te ver hoeven terug te denken.

3. Wees alert op contextuele factoren die de meting kunnen beïnvloeden, los van het te monitoren behandel-effect. Bij klinische uitkomsten kunnen specifieke factoren een rol spelen, zoals ontstekingsmarkers (bijv. CRP) die verhoogd zijn door een intercurrente infectie in plaats van de te monitoren aandoening. Bij patiëntgerapporteerde uitkomsten kunnen niet-behandeling-gerelateerde factoren de scores beïnvloeden, zoals stressvolle levensgebeurtenissen of seizoenseffecten. In het algemeen geldt: hoe specifieker de uitkomst voor de aandoening of behandeling, hoe minder gevoelig voor dergelijke confounding.

Toelichting

Naast betrouwbaarheid (zie 4.4) speelt ook validiteit een rol bij de interpretatie van individuele metingen. Verschillende vormen van vertekening kunnen ertoe leiden dat een gemeten waarde niet volledig overeenkomt met de werkelijke toestand van de patiënt.

Bij patiëntgerapporteerde uitkomsten is recall bias een bekende bron van vertekening. Wanneer patiënten wordt gevraagd terug te denken aan een langere periode (bijvoorbeeld de afgelopen maand), kunnen recente ervaringen zwaarder wegen dan ervaringen aan het begin van die periode. Het kiezen van een korte of geen recallperiode, en het afstemmen van meetmomenten op deze periode, kan dit beperken.

Een verwant fenomeen is de disability paradox: patiënten met ernstige beperkingen rapporteren soms een kwaliteit van leven die hoger is dan zorgprofessionals verwachten. Dit kan leiden tot onderschatting van zorgbehoeften wanneer scores worden vergeleken met externe referenties.

Contextuele factoren kunnen zowel bij patiëntgerapporteerde als bij klinische uitkomsten de meting beïnvloeden, los van het te monitoren behandel-effect. Bij klinische uitkomsten zijn veel markers specifiek: CRP en andere ontstekingsmarkers weerspiegelen ontstekingsactiviteit, maar kunnen ook (tijdelijk) verhoogd zijn door een intercurrente infectie of recente inspanning. Bij patiëntgerapporteerde uitkomsten kunnen niet-behandeling-gerelateerde factoren de scores beïnvloeden, zoals stressvolle levensgebeurtenissen, seizoenseffecten of veranderingen in de thuissituatie. In het algemeen geldt: hoe specifieker de uitkomst voor de aandoening of behandeling, hoe minder gevoelig voor dergelijke confounding.

Een belangrijke nuance is dat consistente vertekening bij longitudinale metingen minder problematisch is wanneer het primair gaat om het beloop: als dezelfde bias bij elke meting optreedt, blijft de verandering over tijd interpreteerbaar. Dit geldt echter niet wanneer de gemeten waarde wordt vergeleken met afkapwaarden of referentiegroepen. De vertekening kan dan wel degelijk tot verkeerde conclusies leiden.

4.4 Betrouwbaarheid van individuele metingen

Is een gemeten waarde op één moment betrouwbaar genoeg om te gebruiken in het gesprek?

Aanbevelingen

1. Gebruik uitkomstmaten waarvoor aannemelijk is dat individuele metingen voldoende precies zijn voor individueel gebruik.
2. Patiëntgerapporteerde vragenlijsten die bestaan uit meerdere items die samen één onderliggend construct meten? Hanteer een betrouwbaarheidscoëfficiënt van circa 0,90 of hoger. Bepaal deze bij voorkeur in de eigen populatie.
3. Uitkomsten die professioneel oordeel vereisen, zoals klinische scores of beeldvormingsbeoordelingen? Beoordeel de overeenstemming tussen beoordelaars (inter-rater) en binnen dezelfde beoordelaar over tijd (intra-rater). Hanteer hiervoor bij voorkeur een intraclass correlatiecoëfficiënt (ICC) van 0,90 of hoger.
4. Klinische of samengestelde uitkomsten waarvoor geen betrouwbaarheidscoëfficiënt beschikbaar is? Beoordeel de bronnen van meetfout die in deze context relevant zijn, zoals fouten bij handmatige extractie uit vrije tekstvelden of variatie door verschillen in meetprocedure.
5. Wees terughoudend met single-item uitkomsten, met name wanneer deze een beperkt aantal antwoordcategorieën hebben. Vermijd conclusies op basis van één meting.
6. Wanneer een instrument de aanbevolen betrouwbaarheidsdrempel niet haalt, wees dan expliciet terughoudend bij de interpretatie van individuele scores. Overweeg of het instrument toch kan worden ingezet met aanvullende duiding over de onzekerheid, of dat een betrouwbaarder alternatief beschikbaar is.
7. Maak een bewuste keuze tussen het tonen van individuele itemscores, domeinscores of samengestelde (composite) scores, afhankelijk van het gespreksdoel. Individuele items geven gedetailleerde informatie maar zijn gevoeliger voor meetfout; samengestelde scores zijn stabiel maar minder specifiek.

Toelichting

Bij individueel gebruik heeft een meetfout een groter verstorend effect dan bij analyses op groepsniveau, omdat toevallige meetruis zich bij één patiënt niet kan uitmiddelen. Daarom is voor interpretatie van een gemeten waarde op één moment een hoge mate van betrouwbaarheid vereist.

Voor patiëntgerapporteerde vragenlijsten die bestaan uit meerdere items die samen één onderliggend construct meten, kan deze betrouwbaarheid worden samengevat in een betrouwbaarheidscoëfficiënt (zoals Cronbach's alfa of verwantematen). Zo'n coëfficiënt geeft een indicatie van de mate waarin individuele scores vrij zijn van meetfout. Voor individueel gebruik wordt een hoge waarde aanbevolen (circa 0,90 of hoger).

Voor uitkomsten die professioneel oordeel vereisen, zoals klinische beoordelingschalen of classificaties op basis van beeldvorming, is de overeenstemming tussen en binnen beoordelaars een belangrijke indicator van betrouwbaarheid. De intraclass correlatiecoëfficiënt (ICC) kwantificeert deze overeenstemming. Voor individueel gebruik wordt een ICC van 0,90 of hoger aanbevolen.

Voor andere typen uitkomsten, zoals klinische indices of samengestelde maten, is een dergelijke samenvattende maat niet altijd beschikbaar. In deze gevallen moet betrouwbaarheid worden beoordeeld aan de hand van potentiële bronnen van meetfout. Uitkomsten die handmatig worden geëxtraheerd uit vrije tekstvelden zijn bijvoorbeeld gevoeliger voor menselijke fouten dan labwaarden die automatisch worden uitgelezen. Dergelijke overwegingen vragen om een expliciete en zorgvuldige afweging door de gebruiker.

4.5 Responsiviteit: vermogen om verandering te meten

Kan de gebruikte uitkomstmaat veranderingen in de gezondheidstoestand meten die in dit zorgtraject verwacht worden?

Aanbevelingen

1. Breng vooraf in kaart welk type verandering in dit zorgtraject verwacht wordt (richting, timing en grootteorde).
2. Ga na of er bewijs is dat veranderingen in scores op het instrument samenhangen met veranderingen in de klinische toestand van de patiënt.
3. Beoordeel in de eigen populatie hoe goed het instrument de verwachte veranderingen kan meten. Bekijk hiervoor de verdeling van scores op de verschillende meetmomenten en ga na of er vloer- of plafondeffecten optreden. Haalt een substantieel deel van de patiënten bij aanvang of halverwege het traject al de minimale of maximale score? Dan is er bij die patiënten onvoldoende ruimte om verandering te detecteren.

4. Als er meerdere instrumenten beschikbaar zijn, geef de voorkeur aan het instrument met de minste vloer- en plafondeffecten in de eigen populatie.
5. Wanneer onderbouwing voor responsiviteit ontbreekt, wees terughoudend met conclusies op basis van (het uitblijven van) verandering en maak deze onzekerheid expliciet in het gesprek.

Toelichting

In N=1-toepassingen vormt verandering vaak de aanleiding voor gesprek of bijsturing. Daarom is het essentieel dat een instrument veranderingen kan laten zien die passen bij het zorgtraject, zowel qua richting als qua omvang. Dit vereist ten eerste dat het instrument in het algemeen gevoelig is voor veranderingen in de klinische toestand (responsiviteit). In de tweede plaats moet het instrument in de specifieke populatie voldoende ruimte bieden om de verwachte veranderingen zichtbaar te maken.

Vloer- en plafondeffecten beperken deze ruimte. Wanneer een groot deel van de patiënten bij aanvang al de laagste score heeft (vloereffect), kan verslechtering niet worden gedetecteerd. Wanneer veel patiënten halverwege het traject al de hoogste score bereiken (plafondeffect), is verdere verbetering niet zichtbaar. Het vergelijken van scoreverdelingen op verschillende meetmomenten geeft inzicht in deze effecten.

Als dergelijke informatie ontbreekt, vraagt het gebruik van het instrument om extra voorzichtigheid bij interpretatie en om expliciete reflectie op de vraag of het uitblijven van verandering daadwerkelijk betekenisvol is.

4.6 Meetfout en betrouwbaarheid van verandering

Welke verandering in de uitkomst kan worden onderscheiden van meetfout in deze populatie en zorgcontext?

Aanbevelingen

1. Bepaal bij voorkeur in de eigen populatie hoe groot de minimale verandering is die kan worden onderscheiden van meetfout, bijvoorbeeld in termen van een minimale detecteerbare verandering (SDC).
2. Gebruik deze informatie om veranderingen op individueel niveau te interpreteren en om te bepalen of een waargenomen verandering aanleiding geeft tot een gesprek of bijstelling van beleid.

3. Wees alert op grillige patronen in grafieken: wanneer scores sterk fluctueren tussen meetmomenten zonder duidelijke klinische verklaring, kan dit wijzen op een hoge meetfout. Overweeg in dat geval het beloop over meerdere meetmomenten te middelen of een betrouwbaarder instrument te kiezen.
4. Wanneer deze informatie niet beschikbaar is, maak gebruik van gegevens uit de literatuur over meetfout, test-hertestbetrouwbaarheid of SDC in vergelijkbare populaties en zorgcontexten.
5. Ontbreekt ook deze onderbouwing, wees dan expliciet terughoudend in de interpretatie van kleine veranderingen. Baseer conclusies bij voorkeur op patronen over meerdere meetmomenten.

Toelichting

Elke uitkomstmaat kent meetfout, waardoor herhaalde metingen bij een onveranderde patiënt variatie laten zien. De minimale detecteerbare verandering (smallest detectable change, SDC) geeft aan hoe groot een verandering moet zijn voordat deze met redelijke zekerheid kan worden onderscheiden van meetfout. Deze parameter is daarom essentieel voor het interpreteren van verandering op individueel niveau.

De SDC kan worden berekend als $SDC = 1,96 * \text{wortel}(2) * SEM$, waarbij SEM de standaardmeetfout is. De SEM hangt direct samen met de betrouwbaarheid van het instrument: hoe hoger de betrouwbaarheid (zie 4.4), hoe kleiner de SEM en dus de SDC. Dit betekent dat instrumenten met een hoge betrouwbaarheid dus kleinere veranderingen betrouwbaar kunnen detecteren.

Instrumenten met een lage betrouwbaarheid en dus een relatief hoge SDC laten in de praktijk vaak een grillig patroon zien in grafieken, ook wanneer de klinische toestand van de patiënt stabiel is. Een dergelijk zaagpatroon maakt het lastig om in de spreekkamer een zinvol gesprek te voeren over hoe de toestand van de patiënt verandert. Is de score nu echt gedaald, of is het meetruis? Dit kan onnodig alarm veroorzaken of het vertrouwen in de meting ondermijnen. Mogelijke oplossingen zijn het middelen van meerdere metingen om de ruis te dempen, of het overstappen op een instrument met hogere betrouwbaarheid.

Voor patiëntenvragenlijsten waarvan de items samen één construct vormen, kan gebruik worden gemaakt van item response theorie (IRT) in combinatie met populatie- of empirische priors om een stabiel beeld te geven. Bij deze benadering wordt de geschatte score van een individuele patiënt gecombineerd met voorkennis over de scoreverdeling in de populatie, waardoor extreme metingen worden gedempt. PROMIS-instrumenten maken bijvoorbeeld gebruik van deze methodiek.

Idealiter wordt de SDC bepaald in de populatie en zorgcontext waarin de uitkomstmaat wordt toegepast, omdat meetfout kan verschillen tussen settings en patiëntgroepen. Wanneer dergelijke lokale informatie ontbreekt, kunnen schattingen uit de literatuur houvast bieden, mits de populatie en meetopzet voldoende vergelijkbaar zijn.

Zonder inzicht in de omvang van meetfout bestaat het risico dat willekeurige fluctuaties worden geïnterpreteerd als klinisch relevante verandering. In zulke situaties is het belangrijk om veranderingen altijd in context te plaatsen, onzekerheid expliciet te benoemen en het beloop over meerdere meetmomenten te volgen.

In de spreekkamer betekent dit concreet: leg aan de patiënt uit dat een kleine score-verandering niet altijd betekenisvol is. Toon bij voorkeur het patroon over meerdere metingen in plaats van één verschil te benoemen.

4.7 Interpretatie van ernst, drempelwaarden en relevante verandering

Welke gemeten waarden of veranderingen in de uitkomst vragen om duiding, actie of heroverweging van het behandelbeleid?

Aanbevelingen

1. Gebruik afkapwaarden en ernstcategorieën uitsluitend voor de duiding van de actuele toestand van de patiënt (cross-sectioneel), zoals ernstindelingen (bijv. mild-matig-ernstig), remissiedrempels of de Patient Acceptable Symptom State (PASS).
2. Baseer afkapwaarden en ernstcategorieën op bestaande, onderbouwde referenties, bij voorkeur uit de literatuur of gevalideerde uitkomstensets. Vermijd het ad hoc vaststellen van drempels.
3. Gebruik vaste klinische relevantiedrempels (zoals MIC/MCID) niet voor de interpretatie van individuele veranderscores. Deze drempels zijn ontwikkeld op groepsniveau en houden geen rekening met individuele meetfout (zie 4.6).
4. Duid veranderingen bij individuele patiënten primair in relatie tot meetfout en verwachte variabiliteit (zie 4.6), en niet door vergelijking met gemiddelde veranderscores uit populaties.
5. Maak expliciet of een uitkomstmaat primair bedoeld is voor het duiden van een toestand, het volgen van verbetering, het signaleren van verslechtering, of een combinatie daarvan, en stem de interpretatie en visualisatie hierop af.

Toelichting

Bij het interpreteren van N=1-uitkomstinformatie is het essentieel onderscheid te maken tussen de duiding van de actuele toestand van een patiënt en de interpretatie van verandering over tijd. Afkapwaarden en ernstcategorieën zijn bij uitstek geschikt voor het eerste doel. Zij helpen om een gemeten score te plaatsen in termen van ernst, controle of acceptabele symptomatologie, zoals bij remissie-indicatoren of de Patient Acceptable Symptom State (PASS).

Voor deze cross-sectionele duiding is het van belang gebruik te maken van bestaande, inhoudelijk onderbouwde drempelwaarden. Deze zijn doorgaans ontwikkeld op basis van klinische expertise en empirisch onderzoek. Ze bieden een herkenbaar en consistent referentiekader voor het gesprek in de spreekkamer. Het gebruik van willekeurige of lokaal vastgestelde drempels vergroot het risico op inconsistente interpretatie. Een aanvullende methode om de actuele toestand te duiden is de vergelijking met andere patiënten: hoe scoort deze patiënt ten opzichte van vergelijkbare anderen? Dit kan helpen de score in perspectief te plaatsen.

Voor veranderingen binnen een individuele patiënt geldt een andere logica. Klinische relevantiedrempels zoals Minimally Important Change (MIC) of Minimally Clinically Important Change (MCID) zijn afgeleid op groepsniveau en houden geen rekening met individuele meetfout. Een verandering bij een individuele patiënt is alleen betekenisvol als deze groter is dan de meetfout die verwacht wordt bij een stabiele patiënt. Gebruik hiervoor de minimale detecteerbare verandering (SDC) als referentie, niet de gemiddelde verandering uit populatiestudies. Toepassing van MIC/MCID op individuele veranderscores kan daardoor leiden tot overinterpretatie (te stellige conclusies op basis van een verschil dat binnen de meetfout valt) en wordt afgeraden.

Door dit onderscheid expliciet te maken, kan uitkomstinformatie zowel cross-sectioneel als longitudinaal worden gebruikt op een methodologisch verantwoorde manier die het gesprek ondersteunt, zonder schijnzekerheid te creëren.



4.8 Begrijpelijke en doelgerichte presentatie van uitkomstinformatie

Is de uitkomstinformatie zó ontworpen dat zorgprofessional en patiënt deze correct, snel en zonder systematische misinterpretatie kunnen begrijpen?

Aanbevelingen

1. Gebruik eenvoudige, gangbare grafiektypen (lijn- of staafdiagrammen) en vermijd complexe of decoratieve visualisaties (zoals radar/spiderdiagrammen).
2. Voorzie elke visualisatie van een duidelijke titel en, waar nodig, een korte toelichting, waarin expliciet wordt gemaakt wat wordt getoond en met welk doel.
3. Combineer visuele weergave altijd met numerieke waarden, zodat interpretatie niet uitsluitend berust op visuele schatting.
4. Gebruik kleur functioneel en eenduidig, bijvoorbeeld om richting (beter/slechter) of signalering aan te geven, en zorg dat de betekenis van kleuren expliciet wordt uitgelegd en niet alleen impliciet wordt verondersteld.
5. Toon per grafiek één uitkomstdomein, en vermijd het combineren van inhoudelijk verschillende constructen (zoals pijn en stemming) in één figuur.
6. Wees terughoudend met het opdelen van assen in categorieën of zones, omdat met name patiënten dergelijke indelingen vaak verkeerd begrijpen.
7. Gebruik consistente schalen over tijd en tussen metingen, en vermijd het in- of uitzoomen van assen zonder expliciete duiding.
8. Toets ontwerpkeuzes bij de beoogde gebruikers, bijvoorbeeld via scenario-tests of korte evaluaties van begrip, en pas het ontwerp hierop aan.
9. Houd rekening met beperkte gezondheidsvaardigheden: gebruik eenvoudige taal, vermijd jargon in labels en titels, en overweeg pictogrammen of uitlegginnen bij complexe concepten zoals onzekerheid.

Toelichting

De manier waarop men uitkomstinformatie presenteert heeft grote invloed op hoe zij wordt begrepen en gebruikt. Onderzoek naar PROM-dashboards en gezondheidsvisualisaties laat zien dat eenvoudige, vertrouwde grafiekvormen zoals lijn- en staafdiagrammen doorgaans het best worden geïnterpreteerd door zowel patiënten als zorgprofessional. Complexere visualisaties, zoals radar- of spiderdiagrammen, vragen meer cognitieve inspanning en leiden vaker tot interpretatiefouten. Onder andere doordat de betekenis van assen en oppervlakken niet eenduidig is.

Het toevoegen van numerieke waarden aan grafieken vermindert de kans op misinterpretatie en voorkomt dat gebruikers conclusies moeten trekken op basis van visuele indrukken alleen. Tegelijkertijd kan het gebruik van kleur helpen om richting of urgentie snel zichtbaar te maken. Mits de betekenis expliciet wordt gemaakt en kleur niet de enige drager van informatie is. Zonder uitleg kunnen kleuren onbedoeld sturen of verwarren, zeker in stressvolle consultsituaties.

Het combineren van meerdere uitkomstdomeinen in één grafiek suggereert samenhang of vergelijkbaarheid die inhoudelijk vaak ontbreekt. Dit vergroot het risico dat verbeteringen en verslechtingen tegen elkaar worden “weggestreept” in het hoofd van de gebruiker. Ook categorische indelingen van continue schalen (bijvoorbeeld zones op de y-as) blijken vaak lastig te interpreteren. Ze kunnen een vals gevoel van precisie of drempelwerking creëren.

Alternatieve visuele formats zoals stoplichtkleuren of symboolreeksen kunnen toegankelijker zijn voor patiënten met beperkte gezondheidsvaardigheden, mits getoetst op correct begrip.

Tot slot is begrijpelijkheid geen theoretisch gegeven maar een empirische vraag. Wat voor ontwikkelaars logisch lijkt, blijkt in de praktijk niet altijd goed te werken. Daarom is het essentieel om dashboards te testen bij de daadwerkelijke gebruikers en ontwerpkeuzes waar nodig aan te passen.

Referenties

- Bantug, E. T., Coles, T., Smith, K. C., Snyder, C. F., Rouette, J., & Brundage, M. D. (2016). Graphical displays of patient-reported outcomes (PRO) for use in clinical practice: what makes a pro picture worth a thousand words? *Patient Education and Counseling*, 99(4), 483-490.
- Brundage, M. D., Smith, K. C., Little, E. A., Bantug, E. T., & Snyder, C. F. (2015). Communicating patient-reported outcome scores using graphic formats: results from a mixed-methods evaluation. *Quality of Life Research*, 24(10), 2457-2472.
- Calvert, M., Kyte, D., Mercieca-Bebber, R., Slade, A., Chan, A. W., King, M. T., ... & SPIRIT-PRO Group. (2018). Guidelines for inclusion of patient-reported outcomes in clinical trial protocols: the SPIRIT-PRO extension. *JAMA*, 319(5), 483-494.
- Dowding, D., Merrill, J. A., Onorato, N., Baez, Y., & Rosati, R. J. (2018). The impact of home care nurses' numeracy and graph literacy on comprehension of visual display information: implications for dashboard design. *Journal of the American Medical Informatics Association*, 25(2), 175-182.
- Engels, N., van der Nat, P. B., Ankersmid, J. W., Prick, J. C. M., Parent, E., The, R., ... & van den Dorpel, M. A. (2022). Development of an online patient decision aid for kidney failure treatment modality decisions. *BMC Nephrology*, 23, 245.
- Few, S. (2006). *Information Dashboard Design: The Effective Visual Communication of Data*. O'Reilly Media.
- Greenhalgh, J., Long, A. F., & Flynn, R. (2005). The use of patient reported outcome measures in routine clinical care: Lack of impact or lack of theory? *Social Science & Medicine*, 60(4), 833-843.
- Greenhalgh, J. (2009). The applications of PROs in clinical practice: What are they, do they work and why? *Quality of Life Research*, 18(1), 115-123.
- ICHOM - International Consortium for Health Outcomes Measurement. <https://www.ichom.org/standard-sets/>
- ISOQOL (2015). *User's Guide to Implementing Patient-Reported Outcomes Assessment in Clinical Practice, Version 2*. International Society for Quality of Life Research.
- Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155-163.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Addison-Wesley.
- Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., ... de Vet, H. C. W. (2010). The COSMIN checklist for evaluating the methodological quality of studies on measurement properties. *Quality of Life Research*, 19(4), 539-549.
- Prinsen, C. A. C., Mokkink, L. B., Bouter, L. M., Alonso, J., Patrick, D. L., de Vet, H. C. W., & Terwee, C. B. (2018). COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of Life Research*, 27(5), 1147-1157.
- Revicki, D. A., Hays, R. D., Cella, D., & Sloan, J. (2008). Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *Journal of Clinical Epidemiology*, 61(2), 102-109.
- Snyder, C. F., Aaronson, N. K., Choucair, A. K., Elliott, T. E., Greenhalgh, J., Halyard, M. Y., ... & Santana, M. (2012). Implementing patient-reported outcomes assessment in clinical practice: a review of the options and considerations. *Quality of Life Research*, 21(8), 1305-1314.
- Snyder, C. F., Smith, K. C., Bantug, E. T., Tolber, R., Blackford, A. L., & Brundage, M. D. (2017). What do these scores mean? Presenting patient-reported outcomes data to patients and clinicians. *Cancer*, 123(10), 1848-1859.
- Stull, D. E., Leidy, N. K., Parasuraman, B., & Chassany, O. (2009). Optimal recall periods for patient-reported outcomes: challenges and potential solutions. *Current Medical Research and Opinion*, 25(4), 929-942.
- Schwarz, N. (2007). Retrospective and concurrent self-reports: The rationale for real-time data capture. In A. A. Stone et al. (Eds.), *The Science of Real-Time Data Capture: Self-Reports in Health Research* (pp. 11-26). Oxford University Press.
- Terwee, C. B., Prinsen, C. A. C., Chiarotto, A., Westerman, M. J., Patrick, D. L., Alonso, J., ...

- de Vet, H. C. W. (2018). COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Quality of Life Research*, 27(5), 1159-1170.
- Tubach, F., Ravaud, P., Baron, G., Falissard, B., Logeart, I., Bellamy, N., ... & Dougados, M. (2005). Evaluation of clinically relevant states in patient reported outcomes in knee and hip osteoarthritis: the patient acceptable symptom state. *Annals of the Rheumatic Diseases*, 64(1), 34-37.
- de Vet, H. C. W., Terwee, C. B., Knol, D. L., & Bouter, L. M. (2006). When to use agreement versus reliability measures. *Journal of Clinical Epidemiology*, 59(10), 1033-1039.
- de Vet, H. C. W., Terwee, C. B., Ostelo, R. W., Beckerman, H., Knol, D. L., & Bouter, L. M. (2006). Minimal changes in health status questionnaires: distinction between minimally detectable change and minimally important change. *Health and Quality of Life Outcomes*, 4, 54.
- de Vet, H. C. W., Terwee, C. B., Mokkink, L. B., & Knol, D. L. (2011). *Measurement in Medicine: A Practical Guide*. Cambridge University Press.
- Van der Wees, P. J., Verkerk, E. W., Verbiest, M. E. A., Zuidgeest, M., Bakker, C., Braspenning, J., ... & Van Dulmen, S. A. (2019). Development of a framework with tools to support the selection and implementation of patient-reported outcome measures. *Journal of Patient-Reported Outcomes*, 3(1), 75.
- Wiering, B., de Boer, D., & Delnoij, D. (2017). Patient involvement in the development of patient-reported outcome measures: a scoping review. *Health Expectations*, 20(1), 11-23.
- Revicki, D., Hays, R. D., Cella, D., & Sloan, J. (2008). Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *Journal of Clinical Epidemiology*, 61(2), 102-109.
- King, M. T. (2011). A point of minimal important difference (MID): a critique of terminology and methods. *Expert Review of Pharmacoeconomics & Outcomes Research*, 11(2), 171-184.
- de Vet, H. C. W., Terwee, C. B., Mokkink, L. B., & Knol, D. L. (2011). *Measurement in Medicine: A Practical Guide*. Cambridge University Press.

5. Hoe verzamel en duid je uitkomstinformatie: 'patients like me'

Inleiding

'Patients like me'-toepassingen gebruiken uitkomstinformatie van groepen patiënten die lijken op de persoon die in de spreekkamer zit. Door kenmerken en context van de individuele patiënt te koppelen aan gegevens van vergelijkbare patiënten, ontstaat een persoonlijkere inschatting van de prognose voor de patiënt, evenals mogelijke behandeluitkomsten voor verschillende behandelopties. Deze 'patients like me'-dashboards kunnen zowel de patiënt als de zorgprofessional ondersteunen bij het proces van samen beslissen.

De aanbevelingen in dit hoofdstuk richten zich primair op ontwikkelaars van 'patients like me'-dashboards en beheerders van uitkomstinformatie. Voor zorgprofessionals vormen zij achtergrondinformatie bij het beoordelen en inzetten van dergelijke toepassingen in de spreekkamer.

Voor verdieping in de methodologische onderbouwing van de beschreven aanbevelingen wordt verwezen naar de methodologische achtergrond bij dit kader.

Binnen deze toepassingen kunnen twee hoofddoelen worden onderscheiden:

1. Het informeren van patiënten over hun te verwachten uitkomst wanneer ze een behandeling hebben gekozen

In deze toepassing ligt de nadruk op prognostische informatie: op basis van de kenmerken van de patiënt én de gekozen behandeling een inschatting maken van het te verwachten beloop of de te verwachten uitkomst. Hierbij gaat het primair om het bieden van realistische verwachtingen over bijvoorbeeld herstel, complicaties of kwaliteit van leven.

Het ontsluiten van uitkomstinformatie voor prognose

De uitkomstinformatie die voor deze toepassing wordt gebruikt kan men op verschillende manieren ontsluiten. In dit kader onderscheiden we de volgende twee methoden:

- Men kan gebruik maken van bijvoorbeeld lokale (d.w.z. van de zorginstelling) of nationale historische data, die inzichtelijk worden gemaakt via dashboards of analysetools (zoals power-BI). Er wordt een groep vergelijkbare patiënten zorgvuldig samengesteld op basis van kenmerken die relevant zijn voor de gezondheid en het te verwachten beloop van de aandoening. Denk hierbij aan leeftijd, geslacht, diagnose, ernst van de ziekte of comorbiditeiten. Het proces waarin men patiënten verdeelt in groepen met vergelijkbare kenmerken, heet *stratificatie*. Door alleen gegevens te gebruiken van patiënten binnen een stratum, krijg je een realistischer beeld van het te verwachten beloop voor een patiënt die op deze groep lijkt. De gemiddelde of mediaan uitkomst in die groep wordt vervolgens teruggekoppeld als referentiepunt voor de huidige patiënt.
- Men kan gebruik maken van een *voorspelmodel*. Dit is een statistisch model dat de kans op een bepaalde uitkomst voor individuele patiënten schat op basis van meerdere kenmerken die zijn afgeleid uit historische patiëntgegevens zoals leeftijd, bloedwaarden, eerdere behandelingen en leefstijlfactoren (de ontwikkeldata). In tegenstelling tot de stratificatiemethode, die werkt met groepsgemiddelden, geeft een voorspelmodel een kans op een uitkomst die specifiek geldt voor één individuele patiënt. Het is daarbij essentieel dat het model is ontwikkeld voor een patiëntenpopulatie waarin de betreffende patiënt valt; anders kan de voorspelling misleidend zijn.

2. Het ondersteunen van patiënten bij het vergelijken van de mogelijke behandelopties. Hiermee krijgen zij inzicht in de mogelijke verschillen tussen behandelingen wat betreft uitkomsten.

In dit geval verschuift de focus van prognose naar vergelijking: niet alleen wat de verwachte uitkomst is, maar hoe deze uitkomst zou verschillen afhankelijk van de gekozen behandeling. Het 'patients like me'-dashboard presenteert dan de verwachte uitkomsten voor meerdere behandelopties naast elkaar, met als doel het ondersteunen van gezamenlijke besluitvorming. Hierbij wordt impliciet een causaal contrast gepresenteerd: wat zou het verschil zijn in uitkomst als dezelfde patiënt behandeling A zou krijgen in plaats van behandeling B? Dit verschil correspondeert met het causale effect van behandeling A ten opzichte van behandeling B.

Het 'patients like me'-dashboard geeft in deze situatie de verwachte uitkomst van verschillende behandelingen (bijvoorbeeld behandeling A en B). Dit representeert een causaal contrast op groepsniveau: het vergelijkt wat er gemiddeld gebeurde in een groep patiënten die behandeling A kreeg met wat er gebeurde in een vergelijkbare groep die behandeling B kreeg. De interpretatie van dit verschil als een causaal effect hangt echter sterk af van de wijze waarop deze groepen tot stand zijn gekomen en van de gebruikte data.



In de literatuur over causale inferentie wordt dit formeel beschreven aan de hand van het concept van potential outcomes. Voor elke individuele patiënt worden twee (of meer) potentiële uitkomsten gedefinieerd: één uitkomst die zou optreden wanneer de patiënt behandeling A krijgt, en één wanneer dezelfde patiënt behandeling B zou krijgen. Het fundamentele probleem van causaliteit is dat we nooit beide potentiële uitkomsten tegelijk kunnen observeren voor dezelfde patiënt. In de praktijk zien we alleen de uitkomst van de daadwerkelijk ontvangen behandeling. De alternatieve uitkomst blijft een hypothetisch scenario. Dit fundamentele probleem maakt dat causale uitspraken alleen onder specifieke aannames mogelijk zijn.

De mate waarin deze aannames gerechtvaardigd zijn, hangt sterk samen met het type data dat men gebruikt voor het ontsluiten van de uitkomst-informatie. In gerandomiseerde studies (trials) wordt door randomisatie van de toewijzing van behandeling verondersteld dat behandelgroepen gemiddeld vergelijkbaar zijn, waardoor verschillen in uitkomst aan de behandeling kunnen worden toegeschreven. In observationele data, die afkomstig zijn uit de klinische praktijk, ontbreekt deze randomisatie. Behandelkeuzes worden daar beïnvloed door patiëntkenmerken, klinische overwegingen en voorkeuren, wat leidt tot confounding. Hierdoor zijn aanvullende aannames en methodologische correcties noodzakelijk om tot een causale interpretatie te kunnen komen (zie 5.2).

Het ontsluiten van uitkomst-informatie voor behandelvergelijkingen

Bij deze toepassing staat niet alleen de verwachte uitkomst centraal, maar vooral het verschil in uitkomst tussen alternatieve behandelingen. Het doel is inzicht te geven in hoe de uitkomst voor een patiënt zou kunnen verschillen wanneer men kiest voor behandeling A, in plaats van behandeling B. In de praktijk is het erg moeilijk om zo'n causale interpretatie te kunnen doen op basis van stratificatie in observationele data. Dit omdat de causale aannames waaraan moet worden voldaan sterk zijn en niet altijd verifieerbaar uit de data. Daarnaast kunnen in zulke observationele ook andere vormen van design-gerelateerde bias zitten (zoals selectie bias, immortal time bias en missende data).

Samenvattend geldt dat 'patients like me'-dashboards zeer geschikt zijn om patiënten te informeren over hun verwachte uitkomst. Voor het eerste doel volstaat het gebruik van observationele data en voorspelmodellen die zijn ontwikkeld op basis van observationele data. Voor het tweede doel, het vergelijken van behandelopties, is echter voorzichtigheid geboden, aangezien dergelijke vergelijkingen impliciet een causale interpretatie veronderstellen.

5.1 Prognose voor de patiënt

5.1.1 Relevantie van de voorspelde uitkomst

Is de voorspelde uitkomst relevant?

Aanbevelingen voor stratificatie én voorspelmodellen

1. Sluit bij het selecteren van uitkomsten in eerste instantie aan bij bestaande kwaliteitsregistraties. Indien niet beschikbaar, kan gekeken worden naar ICHOM-standaardsets. Zorg ervoor dat de resultaten en het moment van afname aansluiten bij de prioriteiten van de patiënt.

Toelichting

Bij het gebruik van voorspelmodellen of stratificatie is het van belang dat de uitkomst die wordt teruggekoppeld en het moment van afname relevant zijn voor de patiënt. Het International Consortium for Health Outcomes Measurement (ICHOM) adviseert om niet één enkele uitkomst te gebruiken, maar een gebalanceerde set die verschillende dimensies van zorgwaarde weerspiegelt. Wanneer een model bijvoorbeeld alleen functioneel herstel na drie maanden voorspelt, terwijl de patiënt vooral pijnvermindering op korte termijn belangrijk vindt, kan aanvullende informatie nodig zijn om tot een goed gesprek te komen. In Nederland wordt primair aangestuurd op het gebruik van uitkomsten die in kwaliteitsregistraties beschikbaar zijn.

5.1.2 Kenmerken van de populatie

Op welke patiënten is het model ontwikkeld?

Aanbevelingen voor stratificatie

1. Maak zichtbaar op het dashboard welke kenmerken zijn gebruikt om de vergelijkbare patiëntengroep samen te stellen.

Toelichting

Bij deze methode koppelt men informatie terug op basis van een zorgvuldig geselecteerde groep patiënten die vergelijkbare kenmerken delen met de huidige patiënt. Deze stratificatie zorgt ervoor dat de vergelijking zo relevant mogelijk is. Factoren zoals leeftijd, geslacht, diagnose, ziektestadium en comorbiditeiten spelen hierbij een belangrijke rol. De stratificatie kan op basis van meerdere variabelen plaatsvinden.

De uitkomst die wordt gepresenteerd is doorgaans een gemiddelde of mediaan van de groep, en dient als referentiepunt voor de patiënt in de spreekkamer. Het is belangrijk dat men transparant maakt hoe deze groep is samengesteld en welke criteria zijn gehanteerd. Wanneer de patiënt aanzienlijk verschilt van de gekozen groep, is de terugkoppeling minder representatief.

Daarnaast is de context van de gegevens van belang: de zorgsetting waarin de vergelijkbare patiënten zijn behandeld moet zoveel mogelijk overeenkomen met die van de huidige patiënt. Dit betekent dat het relevant is om te controleren of de gegevens afkomstig zijn uit een vergelijkbaar type zorginstelling (bijv. academisch vs niet-academisch ziekenhuis) en of het beslismoment vergelijkbaar is.

Aanbevelingen voor voorspelmodellen

1. Een 'patients like me'-dashboard moet informatie bevatten over de groep patiënten die is gebruikt voor de ontwikkeling en validatie van het voorspelmodel.

Toelichting

Bij het beoordelen van een voorspelmodel is het belangrijk om te weten op welke groep patiënten het model oorspronkelijk is ontwikkeld. Een model leert van de gegevens waarop het is ontwikkeld en herkent daarin patronen. Wanneer de patiënt in de spreekkamer verschilt van deze groep, kan de nauwkeurigheid van de voorspelling voor de desbetreffende patiënt afnemen.

Het is van belang om te bekijken in hoeverre de kenmerken van de oorspronkelijke en validatie populatie overeenkomen met die van de patiënt voor wie de voorspelling wordt gebruikt. Daarbij gaat het om factoren zoals leeftijd, geslacht, ziektestadium, comorbiditeiten en eerdere behandelingen. Een model dat is ontwikkeld op gegevens van jonge, relatief gezonde patiënten kan bijvoorbeeld minder goed presteren bij oudere patiënten of mensen met meerdere gezondheidsproblemen. Als men gebruik maakt van een gevalideerd model (zie 5.1.5) is het van belang dit te beoordelen voor de populatie die is gebruikt voor de validatie.

Naast de kenmerken van de patiënten zelf speelt ook de context waarin het model is ontwikkeld een grote rol. De geografische setting en zorgsetting moeten zoveel mogelijk overeenkomen met de setting waarin het model wordt toegepast. Dit betekent dat het relevant is om na te gaan of het model is ontwikkeld in hetzelfde type zorginstelling, of de behandelsetting (nog) vergelijkbaar is en of de timing van het beslismoment overeenkomt.

5.1.3 Type en kwaliteit van de gebruikte data

Welke data wordt gebruikt?

Aanbevelingen voor stratificatie

1. Zorg voor transparantie over de samenstelling van de vergelijkbare patiëntengroep. Gebruik zorgvuldig gekozen variabelen en houd rekening met contextuele factoren, zoals de zorgsetting, bij het definiëren van strata.
2. Houd rekening met de steekproefgrootte van de strata. Hoe meer variabelen men gebruikt voor de stratificatie, hoe kleiner het stratum zal zijn. Hierbij zal men een afweging tussen validiteit (de kans op vertekening) en betrouwbaarheid (de kans op toevalsbevindingen) moeten maken.

Toelichting

Bij terugkoppeling op basis van vergelijkbare patiënten wordt informatie gepresenteerd vanuit een groep patiënten die zoveel mogelijk overeenkomt met de kenmerken van de huidige patiënt. Hierbij maakt men vaak gebruik van observationele data, zoals gegevens uit registraties of elektronische patiëntendossiers. Observationele data weerspiegelen de dagelijkse praktijk, maar kunnen variëren in kwaliteit en volledigheid. Daarom is het zorgvuldig samenstellen van strata essentieel om vertekening te voorkomen.

Een nadeel van stratificatie is dat de beschikbare data wordt opgesplitst in kleinere subgroepen. Naarmate men meer variabelen meeneemt in de stratificatie, wordt de groep patiënten specifiek en daarmee beter vergelijkbaar met de individuele patiënt. Tegelijkertijd leidt dit tot kleinere aantallen per stratum. Hierdoor neemt de statistische onzekerheid toe en worden de uitkomsten gevoeliger voor toevalsvariatie. Daarnaast vergroot stratificatie het risico op multiple testing: doordat analyses in meerdere subgroepen worden uitgevoerd (bijv. man vs vrouw en hoog risico vs laag risico), neemt de kans toe dat men toevallig een significant resultaat vindt. Zonder correctie hiervoor kan dit leiden tot overschatting van effecten en misleidende bevindingen. Het gebruik van voorspelmodellen kan deze limitatie verhelpen.

Aanbevelingen voor voorspelmodellen

1. Zorg voor transparantie over de herkomst en het type data dat is gebruikt voor de ontwikkeling van het voorspelmodel.

Toelichting

Het type data dat voor de ontwikkeling van het voorspelmodel is gebruikt is een belangrijke overweging. Een model kan zijn gebaseerd op routinematig verzamelde zorgdata (bijvoorbeeld uit elektronische patiëntendossiers), op gegevens die verzameld zijn voor onderzoek, bijvoorbeeld klinische trials of grote registraties. Deze databronnen verschillen in kwaliteit, volledigheid en representativiteit. Dat heeft invloed op de toepasbaarheid van het model in een nieuwe situatie.

Veel modellen worden ontwikkeld op basis van data uit elektronische patiëntendossiers. Deze hebben als voordeel dat ze een grote en gevarieerde patiëntenpopulatie bevatten, waardoor de uitkomsten vaak goed aansluiten bij de dagelijkse praktijk. Tegelijkertijd kunnen deze gegevens onvolledig zijn of variëren in kwaliteit, omdat ze oorspronkelijk niet bedoeld zijn voor onderzoek. Modellen die zijn ontwikkeld op basis van klinische trials gebruiken data van patiënten die moesten voldoen aan strikte inclusiecriteria, bijvoorbeeld een beperkte leeftijdscategorie of geen bijkomende aandoeningen. Deze gegevens zijn meestal heel volledig en van hoge kwaliteit, maar de populatie is vaak selectiever. Hierdoor kan de voorspellende kracht van het model lager zijn wanneer het wordt toegepast op patiënten die afwijken van de trialpopulatie, zoals oudere patiënten of patiënten met meerdere aandoeningen.

5.1.4 Consistentie van meting

Zijn de voorspellers en uitkomsten op dezelfde manier gemeten?

Aanbevelingen voor stratificatie en voorspelmodellen

1. Beoordeel of de uitkomsten en voorspellers in de huidige praktijk op dezelfde manier worden gemeten als bij de ontwikkeling van het model. Wanneer dit niet het geval is, is de voorspellende waarde mogelijk verminderd.
2. Maak alleen gebruik van voorspelmodellen waarbij duidelijke definities voor uitkomsten en voorspellers worden gegeven.
3. Maak alleen gebruik van voorspelmodellen die alleen gebruik maken van voorspellers die beschikbaar zijn op het moment van de (behandel) beslissing. Data die ontstaat na behandeling (bijv. postoperatieve scores of complicaties) kunnen geen voorspellers zijn.

Toelichting

Voorspellers zijn de patiëntkenmerken die in het voorspelmodel worden gebruikt om een uitkomst te voorspellen. Het is belangrijk dat men zowel deze voorspellers als de patiëntuitkomsten op dezelfde wijze meet als in de dataset waarop het model is ontwikkeld. Verschillen in meetinstrumenten, definities of meetmomenten kunnen ertoe leiden dat dezelfde variabele in de praktijk iets anders representeert dan in het oorspronkelijke model. Zo kan het gebruik van een andere versie van een vragenlijst, een afwijkende schaalindeling of een ander tijdstip van afname leiden tot systematische verschillen in gemeten waarden en daarmee tot verminderde voorspelprestaties.

Een veelgemaakte fout is het gebruik van postoperatieve complicaties of scores als voorspeller in een model dat diezelfde uitkomsten wil voorspellen. Dit is circulair: dergelijke gegevens ontstaan ná de behandelbeslissing en zijn

op het moment van beslissen nog niet beschikbaar. Bovendien zijn complicaties en postoperatieve scores vaak juist de uitkomsten die het model beoogt te voorspellen. Ze kunnen daarom per definitie geen voorspellers zijn. Een belangrijk uitgangspunt is daarom dat voorspellers uitsluitend mogen bestaan uit informatie die beschikbaar is op het moment van de behandelbeslissing. Alleen wanneer aan deze voorwaarde wordt voldaan, kan de prestatie van het model op een valide en zinvolle manier worden geïnterpreteerd en toegepast binnen de klinische besluitvorming.

5.1.5 Externe validatie

Is het model gevalideerd?

Aanbevelingen voor voorspelmodellen

1. Gebruik bij voorkeur voorspelmodellen die extern zijn gevalideerd in een populatie die vergelijkbaar is met de patiëntenpopulatie waarin het model in de spreekkamer wordt toegepast. Zo is aannemelijk dat de prestaties ook buiten de oorspronkelijke ontwikkelpopulatie behouden blijven.
2. Beoordeel of het voorspelmodel voldoende actueel is en aansluit bij de huidige klinische praktijk. Als het model is gebaseerd op verouderde data of behandelstrategieën, kan men overwegen het model opnieuw extern te valideren (updaten).
3. Kies een voorspelmodel niet alleen op basis van technische prestaties, maar ook op begrijpelijkheid, reproduceerbaarheid en praktische toepasbaarheid. Zo kunnen clinici het model vertrouwen en gebruiken.

Toelichting

Een voorspelmodel moet niet alleen goed presteren in de groep patiënten waarop het is ontwikkeld, maar ook in andere groepen patiënten. Pas wanneer het model ook is getest in een andere populatie, spreken we van een extern gevalideerd model. Hoe consistent de prestaties van het model zijn in verschillende contexten, hoe groter de kans dat de voorspellingen valide zijn voor de patiënt in de spreekkamer. Daarom is een externe validatie essentieel om te beoordelen of het model ook betrouwbaar werkt bij andere patiënten, in andere ziekenhuizen of in een andere zorgsetting.

Tijdens de ontwikkeling kan een model echter patronen leren die specifiek zijn voor de ontwikkeldataset (overfitting). Dit leidt tot zogenoemd optimisme: het model presteert beter in de ontwikkeldata dan in nieuwe, onafhankelijke data. Dit ontstaat doordat het model niet alleen het onderliggende signaal leert, maar ook ruis en toevallige patronen in de ontwikkeldata die niet generaliseerbaar zijn. Als gevolg daarvan worden prestatiematen zoals discriminatie (zie 5.1.7) en kalibratie (zie 5.1.8) vaak te positief ingeschat.

De mate van optimisme hangt samen met de complexiteit van het model en de hoeveelheid beschikbare data. Wanneer het aantal voorspellers groot is ten opzichte van het aantal waargenomen gebeurtenissen, neemt het risico op overfitting toe. Dit laat zien dat complexere modellen niet per definitie beter zijn. Integendeel: uit meerdere systematische reviews blijkt dat geavanceerde machine learning modellen in veel klinische toepassingen geen betere prestaties leveren dan traditionele statistische modellen, zoals logistische regressie.² De toegevoegde waarde van complexere methoden lijkt vooral aanwezig bij zeer grote datasets met complexe datastructuren, zoals beeldvorming, genetische informatie of vrije tekst.

Daarbij komt dat complexere modellen vaak minder transparant zijn. Terwijl traditionele statistische modellen doorgaans goed uitlegbaar zijn, kunnen meer geavanceerde algoritmen functioneren als een ‘black box’, wat het begrip en vertrouwen van medici kan verminderen. De keuze voor een bepaald type model moet daarom niet alleen worden gebaseerd op technische prestaties, maar ook op begrijpelijkheid, reproduceerbaarheid en toepasbaarheid in de praktijk.

Ook wanneer een voorspelmodel voldoet aan gangbare eisen voor modelontwikkeling, blijft externe validatie essentieel. Modellen kunnen in andere populaties, zorgcontexten of tijdsperioden anders presteren door variatie in patiëntkenmerken, diagnostiek of behandelpraktijken. Wanneer blijkt dat de modelprestaties afnemen, betekent dit niet dat het model onbruikbaar is. In veel gevallen kan het model opnieuw worden gevalideerd (updaten) met recentere of meer representatieve data.

5.1.6 Onzekerheid rond de voorspelling

Hoe betrouwbaar is de voorspelling?

Aanbevelingen voor voorspelmodellen

1. Gebruik voorspelmodellen die de onzekerheid van een puntschatting laten zien.

Toelichting

Elke voorspelling die een model geeft, gaat altijd gepaard met een zekere mate van onzekerheid. Een model kan nooit met absolute zekerheid aangeven wat er bij een individuele patiënt zal gebeuren. Daarom worden voorspellingen vaak aangevuld met een betrouwbaarheidsinterval of een andere maat voor onzekerheid. Deze informatie laat zien binnen welk bereik de werkelijke uitkomst waarschijnlijk zal liggen. Door niet alleen naar één getal te kijken, maar ook naar de spreiding daaromheen, ontstaat een beter beeld van hoe precies of hoe onzeker de voorspelling eigenlijk is.

Wanneer het betrouwbaarheidsinterval rond de voorspelling breed is, betekent dit dat het model minder nauwkeurig kan inschatten wat er waarschijnlijk gaat gebeuren. Dit is belangrijke informatie voor het gesprek met de patiënt, omdat het helpt verwachtingen realistisch te houden. Een smal betrouwbaarheidsinterval wijst juist op een meer precieze schatting.

Het is daarom belangrijk dat modellen niet alleen een puntschatting tonen, maar ook duidelijk maken hoeveel onzekerheid daarbij hoort. Een voorspelling zonder betrouwbaarheids- of predictie-interval kan misleidend zijn. Het suggereert dat de uitkomst exact vaststaat, terwijl dat in de praktijk niet het geval is.

5.1.7 Discriminatievermogen

Hoe goed kan dit model onderscheid maken tussen patiënten met verschillende uitkomsten?

Aanbevelingen voor voorspelmodellen

1. Beoordeel de prestatie van de (extern gevalideerde) voorspelmodellen.

Toelichting

Een bruikbaar voorspelmodel moet goed kunnen onderscheiden tussen patiënten die waarschijnlijk een gunstige uitkomst hebben en patiënten met een minder gunstige uitkomst. Dit vermogen om verschillen tussen patiënten correct te herkennen wordt ook wel het discriminatievermogen van het model genoemd. Een model met een goed discriminatievermogen rangschikt patiënten op een manier die overeenkomt met hun daadwerkelijke risico's: patiënten bij wie de uitkomst optreedt krijgen een hogere voorspelde kans, en patiënten bij wie de uitkomst niet optreedt krijgen een lagere voorspelde kans. Voor samen beslissen is vooral relevant of het model op individueel niveau helpt om betekenisvolle verschillen zichtbaar te maken. Wanneer het discriminatievermogen laag is, kan de voorspelling te weinig houvast bieden om keuzes over behandeling te ondersteunen.

Voor binaire uitkomsten wordt dit onderscheid samengevat in statistieken zoals de C-statistiek (ook wel de area-under-the-ROC-curve, AUC, genoemd). Deze waarden liggen tussen de 0,5 en 1. Een waarde van 0,5 betekent dat het model niet beter presteert dan toeval; dit is vergelijkbaar met het opgooien van een munt. Een waarde van 1 discrimineert perfect. Een waarde van 0,7 of hoger wordt meestal gezien als een goede indicatie dat het model patiënten effectief weet te onderscheiden op basis van hun risico.

Voor continue uitkomsten wordt het onderscheidend vermogen van een model meestal uitgedrukt met behulp van de verklaarde variantie, aangeduid als de R^2 . Deze maat ligt tussen 0% en 100% en geeft aan welk deel van de variatie in de uitkomst door het model wordt verklaard. Een R^2 van 0% betekent dat het model de variatie in uitkomst niet kan verklaren, terwijl een R^2 van 100% aangeeft dat het model de uitkomst volledig kan verklaren. Hoe hoger de R^2 , hoe groter het aandeel van de variatie dat door het model wordt verklaard en hoe beter het model verschillen tussen patiënten kan beschrijven.

5.1.8 Kalibratie

Komt de voorspelde kans van het model overeen met wat er in de praktijk gebeurt bij vergelijkbare patiënten?

Aanbevelingen voor voorspelmodellen

1. Gebruik alleen voorspelmodellen waarvoor de kalibratie is beoordeeld.
2. Kalibreer het voorspelmodel op de populatie van toepassing. Binnen een zorginstelling kan dit bijvoorbeeld gaan om lokale historische data.

Toelichting

De kalibratie van een model geeft aan in hoeverre de voorspelde kansen overeenkomen met de werkelijke uitkomsten in de praktijk. Een goed gekalibreerd model voorspelt bijvoorbeeld dat van de patiënten met een geschatte kans van 20% op een bepaalde uitkomst, ongeveer één op de vijf die uitkomst ook daadwerkelijk zal ervaren. Wanneer een model structureel te hoge kansen voorspelt, kan dit patiënten onnodig ongerust maken. Als het juist te lage kansen geeft, bestaat het risico dat patiënten te veel worden gerustgesteld. Een slechte aansluiting tussen voorspelling en werkelijkheid kan daardoor leiden tot beslissingen die niet volledig op betrouwbare informatie zijn gebaseerd.

Het beoordelen van kalibratie gebeurt doorgaans aan de hand van kalibratieplots en statistische maten zoals het kalibratie-intercept en de -helling. Deze geven inzicht in respectievelijk systematische over- of onderschatting en in de mate waarin voorspellingen te extreem of juist te terughoudend zijn. Een goed gekalibreerd model heeft een intercept rond nul en een helling dicht bij één.

Belangrijk is dat kalibratie altijd context- en populatieafhankelijk is. Een model dat goed gekalibreerd is in de populatie waarin het is ontwikkeld, hoeft dat niet automatisch te zijn in een andere setting, zoals een ander ziekenhuis, een andere patiëntengroep of een latere periode. Veranderingen in patiëntkenmerken, behandelpraktijken of meetmethoden kunnen ertoe leiden dat de voorspelde risico's niet langer overeenkomen met de werkelijke uitkomsten. Daarom is het essentieel om de kalibratie te beoordelen in de specifieke populatie waarin het model wordt toegepast en deze zo nodig aan te passen.

5.1.9 Klinische bruikbaarheid

Wanneer kunnen we het model gebruiken in de klinische praktijk?

Hoewel veel voorspelmodellen worden ontwikkeld en gevalideerd, worden slechts weinig modellen daadwerkelijk toegepast in de klinische praktijk. Studies laten zien dat slechts een klein deel van de gepubliceerde modellen uiteindelijk wordt geïmplementeerd in de praktijk. Daarnaast is het aantal modellen dat aantoonbaar leidt tot betere klinische besluitvorming zeer beperkt. Dit laat zien dat goede modelprestaties alleen onvoldoende zijn om klinische bruikbaarheid te garanderen.

Prestatiematen zoals discriminatie en kalibratie geven inzicht in de statistische kwaliteit van een model, maar zeggen op zichzelf weinig over de klinische relevantie. Een model kan bijvoorbeeld goed onderscheid maken tussen hoog- en laagrisicopatiënten, zonder dat dit leidt tot andere of betere behandelbeslissingen. Om die reden is aanvullende evaluatie nodig, bijvoorbeeld via decision curve analysis (DCA). DCA evalueert het netto klinische voordeel van een model over een reeks drempelwaarden waarboven een interventie wordt overwogen. Hiermee kan worden beoordeeld of het gebruik van een model daadwerkelijk leidt tot betere beslissingen dan alternatieve strategieën, zoals iedereen behandelen of niemand behandelen.

Wanneer een voorspelmodel bedoeld is om klinische beslissingen te ondersteunen, kan het onder de Europese regelgeving voor medische hulpmiddelen (MDR) vallen en CE-gemarkeerd moeten worden. Hierbij wordt onder meer beoordeeld of het model veilig is, of het doet wat het belooft en of men het op verantwoorde wijze kan toepassen in de klinische praktijk. Het onderscheid tussen stratificatie en daadwerkelijke voorspelmodellen is hierbij relevant. Modellen die slechts patiënten indelen in risicogroepen op basis van vooraf vastgestelde criteria kunnen soms buiten de strengste regelgeving vallen. Anderzijds worden modellen die individuele voorspellingen doen en daarmee behandelbeslissingen beïnvloeden eerder als medisch hulpmiddel aangemerkt.

5.2 Keuze tussen behandelingen

5.2.1 Causaliteit van behandel-effecten

Hoe zeker is het dat de gevonden verschillen in uitkomsten echt door de behandeling komen?

Aanbevelingen

1. Maak duidelijk of een model bedoeld is om uitkomsten te voorspellen (prognose) of om behandelkeuzes te ondersteunen (causale inferentie). Vermijd het gebruik van voorspelmodellen voor causale beslissingen wanneer zij daar niet voor zijn ontworpen.
2. Gebruik bij voorkeur modellen gebaseerd op trial data wanneer causale interpretatie gewenst is.
3. Wees terughoudend met causale interpretaties bij gebruik van voorspelmodellen die zijn ontwikkeld op basis van observationele data.

Toelichting

Bij het gebruik van voorspelmodellen om behandelbeslissingen te ondersteunen, is het van belang onderscheid te maken tussen modellen die zijn gebaseerd op trial data - ook wel data met gerandomiseerde toewijzing van behandeling genoemd - en modellen die zijn ontwikkeld op basis van observationele data. De mate waarin uitkomsten causale betekenis hebben, hangt sterk af van dit onderscheid.

Wanneer beschikbaar, verdienen voorspelmodellen die zijn gebaseerd op gerandomiseerde klinische studies de voorkeur. In deze data is de toewijzing van behandeling onafhankelijk van patiëntkenmerken, waardoor verschillen in uitkomst kunnen worden toegeschreven aan het effect van de behandeling zelf. Modellen die op deze data zijn gebaseerd, bieden daarom de meest betrouwbare schatting van causale behandel-effecten. Een bekend voorbeeld is het PREDICT-model voor borstkanker, waarin resultaten uit gerandomiseerde studies zijn verwerkt om de verwachte effecten van verschillende behandelopties inzichtelijk te maken (<https://breast.predict.cam/>).



In de praktijk wordt echter vaak gebruik gemaakt van voorspelmodellen die zijn ontwikkeld op basis van observationele data, zoals gegevens uit elektronische patiëntendossiers. Deze data sluiten goed aan bij de dagelijkse klinische praktijk en bevatten vaak grote en diverse populaties. Tegelijkertijd brengen zij belangrijke methodologische beperkingen met zich mee. De keuze voor een behandeling is in deze datasets vaak niet willekeurig, maar hangt samen met patiëntkenmerken zoals ziekte-ernst, comorbiditeit en eerdere behandelgeschiedenis. Hierdoor bestaat een aanzienlijk risico op confounding. Hierdoor kunnen waargenomen verschillen in uitkomst ten onrechte worden toegeschreven aan de behandeling, terwijl zij in werkelijkheid (deels) worden veroorzaakt door onderliggende verschillen tussen patiënten.

Om verschillen in uitkomst toch als causale behandel effecten te kunnen interpreteren, zijn drie aannames essentieel:

1. Conditional exchangeability: Zijn alle relevante confounders meegenomen?

Het risico dat ongemeten of onbekende factoren zowel de behandelkeuze als de uiteindelijke uitkomst beïnvloeden blijft bestaan. Dit wordt confounding genoemd en vormt een belangrijke beperking voor het trekken van causale conclusies over het effect van een behandeling. De uitkomsten die in een dashboard worden gepresenteerd, laten daardoor vooral zien wat er is gebeurd binnen vergelijkbare patiëntgroepen. Ze vormen geen bewijs voor wat er zou zijn gebeurd als dezelfde patiënt een andere behandeling had gekregen.

In de causale literatuur wordt de aanname die nodig is om ondanks deze beperking toch causale uitspraken te kunnen doen aangeduid als conditional exchangeability. Deze aanname houdt in dat, gegeven een set waargenomen patiëntkenmerken, de toewijzing aan een behandeling onafhankelijk is van de potentiële uitkomsten. Met andere woorden: nadat is gecorrigeerd voor relevante variabelen, mogen er geen systematische verschillen meer bestaan tussen patiënten die behandelingen A versus B krijgen. In de praktijk betekent dit dat alle factoren die zowel de behandelkeuze als de uitkomst beïnvloeden gemeten en meegenomen moeten zijn. Dit is een niet te verifiëren aanname.

Een praktisch hulpmiddel om te beoordelen of deze aanname plausibel is, is het gebruik van een Directed Acyclic Graph (DAG). Een DAG is een conceptueel model waarin met pijlen wordt weergegeven welke variabelen elkaar beïnvloeden. Door relevante patiëntkenmerken, behandelingen en uitkomsten expliciet in zo'n schema op te nemen, wordt inzichtelijk welke variabelen mogelijk confounding veroorzaken en of er sprake is van ongemeten confounding.

2. Positivity: Is er voldoende overlap tussen behandelgroepen?

De positivity-aanname vereist dat voor elke combinatie van relevante patiëntkenmerken een reële kans bestaat om elke behandeling te ontvangen. Met andere woorden: voor iedere patiënt moet het in principe mogelijk zijn dat hij of zij zowel behandeling A als behandeling B had kunnen krijgen. In de praktijk is dit niet altijd het geval. Sommige patiënten komen uitsluitend in aanmerking voor één specifieke behandeling, bijvoorbeeld vanwege ziekteernst, contra-indicaties of klinische richtlijnen. Wanneer dergelijke situaties optreden, ontbreekt de basis voor een valide vergelijking, omdat er geen vergelijkbare patiënten bestaan die de alternatieve behandeling ontvingen.

Wanneer deze overlap ontbreekt, zijn causale uitspraken niet langer goed onderbouwd. Het model moet dan extrapoleren buiten het bereik van de beschikbare data. Dit vergroot de kans op vertekening en maakt de resulterende schattingen onbetrouwbaar. Het is daarom essentieel om voorafgaand aan de interpretatie van uitkomsten te beoordelen of er voldoende overlap bestaat tussen behandelgroepen, bijvoorbeeld door het inspecteren van verdelingen van variabelen. Zonder deze controle bestaat het risico dat beslisondersteuning gebaseerd is op onvolledige of misleidende informatie.

3. Consistency: Is de behandeling eenduidig gedefinieerd en vergelijkbaar toegepast?

Deze aanname stelt dat de geobserveerde uitkomst bij een patiënt die een bepaalde behandeling ontvangt, overeenkomt met de potentiële uitkomst onder diezelfde behandeling. Dit veronderstelt dat de behandeling eenduidig is gedefinieerd en op vergelijkbare wijze wordt toegepast. Wanneer behandelingen in de praktijk sterk variëren in uitvoering, intensiteit of context, is het moeilijk om eenduidige causale conclusies te trekken. Bijvoorbeeld “een gezonde leefstijl” versus “een ongezonde leefstijl” kan op veel verschillende manieren worden gedefinieerd en is dus niet eenduidig.

Wanneer aan één of meerdere van deze aannames niet wordt voldaan, moeten voorspellingen met voorzichtigheid worden geïnterpreteerd. Het is daarom essentieel om voorafgaand aan het gebruik van uitkomstinformatie bij behandelbeslissingen expliciet te toetsen of deze aannames plausibel zijn. Bijvoorbeeld door het inspecteren van covariaatverdelingen, het analyseren van overlap tussen behandelgroepen en het kritisch beoordelen van de homogeniteit van behandeldefinities. Naast fundamentele aannames is het bij observationele data essentieel om design-gerelateerde bias, zoals immortal time bias, te adresseren. Target trial emulation biedt hiervoor een raamwerk. Door relevante confounders te includeren wordt randomisatie benaderd (conditional exchangeability). Geschikte inclusiecriteria zorgen voor overlap in propensity scores (positivity), terwijl correcte time-zero-definitie bias door toekomst informatie voorkomt. Door behandelstrategieën expliciet te definiëren wordt consistentie gewaarborgd. Literatuurverwijzingen hiervoor zijn opgenomen in de methodologische achtergrond.

Een belangrijk aanvullend aandachtspunt is dat voorspelmodellen die zijn ontwikkeld op observationele data vaak impliciet het bestaande behandelbeleid weerspiegelen. Zij voorspellen daarmee uitkomsten onder de “huidige zorgpraktijk”, niet onder alternatieve behandelstrategieën. Dit kan leiden tot zogeheten causal blind spots: situaties waarin een laag voorspeld risico juist het gevolg is van eerder ingezette behandeling, en niet van een intrinsiek gunstig ziektebeloop. Het gebruik van dergelijke modellen voor behandelkeuzes kan dan tot verkeerde conclusies leiden.

5.2.2 Onzekerheid bij behandelvergelijkingen

Worden onzekerheden in de uitkomsten duidelijk weergegeven?

Aanbevelingen voor voorspelmodellen

1. Rapporteer onzekerheid rond individuele voorspellingen.
2. Maak onderscheid tussen voorspellende nauwkeurigheid en causale interpretatie.
3. Beschrijf de herkomst en kwaliteit van de data waarop het model is getraind en gevalideerd.

Toelichting

Bij voorspelmodellen is het van belang dat onzekerheid rond individuele voorspellingen expliciet wordt weergegeven. Voorspellingen zijn altijd probabilistisch van aard en geven geen zekere uitkomst, maar een geschatte kans gebaseerd op patronen in historische data. Het is daarom essentieel om naast puntschattingen ook onzekerheidsmarges te tonen, bijvoorbeeld in de vorm van betrouwbaarheids- of predictie-intervallen.

Verder is het van belang om expliciet te zijn over de aannames en beperkingen van het model. Onvolledige registratie, meetfouten, selectiebias en veranderingen in zorgpraktijk kunnen allemaal invloed hebben op de betrouwbaarheid van de voorspellingen. Ook dient aandacht te worden besteed aan de generaliseerbaarheid van het model: een model dat goed presteert in de ontwikkelpopulatie is niet per definitie betrouwbaar in andere populaties of zorgcontexten.

5.2.3 Heterogeniteit van behandel-effecten

Geldt het gevonden effect ook voor mijn patiënt, of zijn er subgroepen die anders reageren?

Aanbevelingen voor voorspelmodellen

1. Gebruik het gemiddelde behandelingseffect niet om conclusies te trekken op individueel patiëntniveau, omdat individuele effecten aanzienlijk kunnen verschillen.

Toelichting

Wanneer behandeluitkomsten worden geëvalueerd, wordt meestal het *gemiddeldebehandeleffect* (average treatment effect, ATE) gerapporteerd. Dit beschrijft het gemiddelde verschil in uitkomst tussen behandeling A en behandeling B in de onderzochte populatie. Hoewel dit waardevolle informatie biedt op populatieniveau, weerspiegelt een gemiddeld effect niet noodzakelijk wat geldt voor iedere individuele patiënt. In de praktijk kunnen behandelresponsen sterk variëren tussen patiënten. Deze variatie wordt aangeduid als *heterogene behandel-effecten* (heterogeneous treatment effects, HTE). Factoren zoals leeftijd, geslacht, ziekte-ernst, comorbiditeit of genetische kenmerken kunnen ervoor zorgen dat patiënten verschillend reageren op dezelfde behandeling.

In de context van samen beslissen met uitkomst-informatie is het van belang te erkennen dat een gemiddeld behandel-effect individuele verschillen kan verhullen. Een behandeling die gemiddeld gunstig uitpakt, kan voor sommige patiënten weinig voordeel bieden of zelfs ongunstig uitpakken. Het negeren van deze heterogeniteit kan ertoe leiden dat uitkomst-informatie niet goed aansluit bij de persoonlijke situatie en voorkeuren van de patiënt. Hierdoor worden beslissingen minder goed afgestemd op wat voor die individuele patiënt werkelijk relevant is.

Daarom is het van belang om te onderzoeken of er subgroepen bestaan waarin het effect van de behandeling afwijkt van het gemiddelde. Subgroepanalyses of modellen die rekening houden met patiëntkenmerken kunnen helpen om te beoordelen of het gevonden gemiddelde effect ook relevant is voor een specifieke patiënt. Dit is met name belangrijk wanneer behandelbeslissingen worden genomen voor patiënten die duidelijk afwijken van de gemiddelde studiepopulatie.

Tegelijkertijd vraagt de interpretatie van subgroepanalyses om voorzichtigheid. Subgroepen bevatten vaak relatief weinig patiënten. Hierdoor zijn schattingen onzekerder en neemt het risico op toevallige bevindingen toe. Dit geldt in het bijzonder voor zogenoemde one-variable-at-a-time subgroepanalyses, waarbij het effect van een behandeling afzonderlijk wordt bekeken binnen één kenmerk

(bijvoorbeeld leeftijd of geslacht). Dergelijke analyses houden geen rekening met de samenhang tussen patiëntkenmerken en vergroten de kans op schijnbare verschillen die statistisch niet robuust zijn. Recente methodologische literatuur, waaronder de PATH-richtlijn (Predictive Approaches to Treatment effect Heterogeneity), benadrukt daarom dat het identificeren van heterogene behandelingseffecten bij voorkeur gebeurt via multivariabele, risicogebaseerde benaderingen. Deze benaderingen modelleren gelijktijdig meerdere patiëntkenmerken en bieden een consistent kader om variatie in behandelrespons te onderzoeken. Ze verkleinen het risico op fout-positieve bevindingen en sluiten beter aan bij de klinische realiteit, waarin behandelbeslissingen zelden op één enkele patiëntkarakteristiek worden gebaseerd.

Samengevat biedt het gemiddelde behandel-effect waardevolle informatie op populatieniveau, maar is het onvoldoende om individuele behandelbeslissingen volledig op te baseren. Inzicht in heterogene behandelingseffecten kan helpen om zorg beter af te stemmen op individuele patiënten, mits deze inzichten zorgvuldig worden geïnterpreteerd en onderbouwd.

5.2.4 Begrijpelijke presentatie van behandelvergelijkingen

Worden de uitkomsten en verschillen op een begrijpelijke manier gepresenteerd?

Aanbevelingen

1. Presenteer gemiddelden voor normaal verdeelde uitkomsten, en hanteer medianen met interkwartielbereik bij scheef verdeelde uitkomsten (zoals tijd tot gebeurtenis, opnameduur of aantal complicaties).
2. Door naast een puntschatting ook het betrouwbaarheidsinterval en de richting van het effect te presenteren, wordt duidelijker hoe de uitkomsten over patiënten zijn verdeeld (bijvoorbeeld: “50% van de patiënten ervaart een opnameduur tussen X en Y”).

Toelichting

Elke uitkomst moet men presenteren binnen een expliciete en betekenisvolle context. Voor samen beslissen met uitkomst-informatie vereist dit dat uitkomsten van geen behandeling (bijv. “afwachtend beleid” of “zien en observeren”) naast behandelingsopties worden gepresenteerd, met behulp van vergelijkbare meeteenheden: dezelfde noemer, tijdshorizon en visualisatievorm. Wanneer van toepassing dient ook de optie van niet curatief behandelen of afwachtend beleid expliciet te worden meegenomen als vergelijkingsoptie. Zo krijgen patiënten een volledig beeld van de beschikbare keuzes.

Uitkomsten met een langere tijdshorizon (bijv. langetermijnsterfte of effecten op de kwaliteit van leven) vereisen vaak aanvullende contextualisering om door patiënten te kunnen worden geïnterpreteerd. Deze uitkomsten kunnen bijvoorbeeld worden vergeleken met de totale sterfte, leeftijdsgebonden veranderingen in de kwaliteit van leven of basisrisico's in de bevolking om een zinvolle interpretatie te ondersteunen.

De geschikte tijdshorizon is sterk contextafhankelijk. Voor sommige aandoeningen is het presenteren van uitkomsten op één enkel tijdstip onvoldoende. In plaats daarvan kan een tijdlijn of traject dat laat zien hoe de uitkomsten zich in de loop van de tijd ontwikkelen, nodig zijn (bijv. symptoomlast, functionele achteruitgang, cumulatief risico). Wanneer men meerdere uitkomsten presenteert, moet de gekozen tijdshorizon voor alle uitkomsten identiek zijn om impliciete framing-effecten te vermijden.

Wanneer het causale verband tussen een interventie en een uitkomst niet stevig is vastgesteld, moet deze onzekerheid expliciet worden vermeld. Patiënten gaan er vaak van uit dat er een causaal verband bestaat tussen behandeling en uitkomst, zelfs wanneer het bewijs observationeel of indirect is. Dit kan leiden tot een verkeerde interpretatie van de voordelen en nadelen.

Uitkomsten

Alle uitkomsten moeten op een neutrale, niet-sturende manier worden gepresenteerd. Voor probabilistische uitkomsten omvat dit een dubbele framing, waarbij zowel positieve als negatieve kaders worden gepresenteerd (bijv. overleving en mortaliteit). Geef uitkomsten over het algemeen numeriek weer aangezien verbale frequentielabels alleen (bijv. “zeldzaam”, “vaak voorkomend”) inconsistent worden geïnterpreteerd en sterk contextafhankelijk zijn.

- **Categorische of ordinale uitkomsten** (bijv. overlijden ja/nee, symptoomcategorieën) worden vaak goed begrepen wanneer ze duidelijk gelabeld zijn.
- **Continue uitkomsten** afgeleid van schalen (bijv. kwaliteit van leven, functionele scores) dient men te presenteren samen met het schaalbereik en de richting (bijv. “2,4 op een schaal van 0-5, waarbij hogere scores een beter functioneren aangeven”), met korte verbale ankers, indien nodig.
- **Gestratificeerde probabilistische uitkomsten** worden het meest intuïtief gecommuniceerd met behulp van natuurlijke frequenties (bijv. “50 van de 1000 patiënten zoals u”), die het perspectief van “patiënten zoals ik” benadrukken. Bij het gebruik van frequenties moet men dezelfde noemer gebruiken voor alle opties (bijv. “50 op 1.000” versus “2 op 1.000”), omdat het veranderen van de noemer cognitieve vertekening introduceert.
- **Voorspelmodel** (individuele risicoschattingen) kan men vaak het beste presenteren als percentages, mits de noemer en de referentiepopulatie duidelijk zijn gespecificeerd.

Onzekerheid

- **Stratificatie:** Betrouwbaarheidsintervallen of interkwartielbereik (afhankelijk van of men het gemiddelde of de mediaan gebruikt) als bereik + interpretatie.
- **Voorspelmodel:** Voorspellingsinterval + interpretatie.

Uitkomstspecifieke overwegingen

- Ordinale uitkomsten (bijv. aantal bloedingen): Rapporteer de mediaan en de interkwartielafstand waar relevant.
- Binaire uitkomsten (kans op een gebeurtenis): Rapporteer het absolute risico met betrouwbaarheidsintervallen; vermijd het gebruik van alleen het relatieve risico.
- Continue uitkomsten (bijv. scores voor kwaliteit van leven):
- Uitkomstenovertijd: Het visualiseren van onzekerheid over tijd (bijv. gearceerde betrouwbaarheidsintervallen) kan helpen om variabiliteit en onzekerheid in het verloop van tijd weer te geven.

Onzekerheid in stratificatie versus voorspelmodel:

- Voor geïndividualiseerde voorspelmodel (waarschijnlijkheid van uitkomst X): Vermeld duidelijk dat voorspellingen schattingen zijn gebaseerd op groepsgegevens.
- Rapporteer voorspellingsintervallen, niet alleen puntschattingen.
- Geef aan hoe de precisie afhangt van:
 - Grootte en representativiteit van de onderliggende populatie.
 - Overeenkomst tussen de patiënt en de gegevensbron.
 - Zorgomgeving.

Visualisaties

Visualisaties moeten numerieke informatie ondersteunen, niet vervangen. Categorische of ordinale uitkomsten vereisen vaak geen visualisatie en kunnen verkeerd worden weergegeven als er ongeschikte afbeeldingen worden gebruikt. Wanneer patiënten mogelijk niet bekend zijn met een concept (bijv. medische hulpmiddelen of procedures), kan men een eenvoudig, neutraal pictogram toevoegen.

Waarschijnlijkheden moet men altijd visualiseren om patiënten met een lagere rekenvaardigheid te ondersteunen. Geordende pictogrammenreeksen hebben de voorkeur voor natuurlijke frequenties in gestratificeerde uitkomsten. Gestapelde kolomdiagrammen zijn geschikt voor percentages in persoonlijke voorspellingen. Boomdiagrammen zijn met name nuttig voor voorwaardelijke waarschijnlijkheden (bijv. screeningsresultaten met ware/valse posities en negatieven). Alle visualisaties moeten zowel de teller als de noemer duidelijk weergeven. Voor zeer kleine waarschijnlijkheden (<1%, en vooral <0,01%) kunnen verschillen moeilijk op een zinvolle manier worden gevisualiseerd. In dergelijke gevallen kunnen zorgvuldig gekozen analogieën op bevolkingsniveau (bijvoorbeeld: “van de 18 miljoen mensen in Nederland...”) het begrip bevorderen, mits de analogie neutraal is en duidelijk als illustratief wordt aangeduid.

Transparantie

Patiënten moet men in begrijpelijke taal informeren over:

- Welke variabelen zijn gebruikt voor stratificatie of voorspelling?
- De gegevensbronnen die ten grondslag liggen aan de schattingen.
- De mate van onzekerheid en de kwaliteit van het behandel-effect.
- Moedig patiënten ook expliciet aan en ondersteun ze om, indien gewenst, aanvullende informatie te zoeken (bijvoorbeeld via links als thuisarts).

Referenties

- Arshi B, Wynants L, Rijnhart E, Reeve K, Cowley LE, Smits LJ. Number of Publications on New Clinical Prediction Models: A Bibliometric Review. *JMIR Med Inform*. 2025 Jul 4;13:e62710.
- Austin, Peter C. et al. Testing multiple statistical hypotheses resulted in spurious associations: a study of astrological signs and health. *Journal of Clinical Epidemiology*, Volume 59, Issue 9, 964 - 969
- Berry DC, Raynor DK, Knapp P, Bersellini E. Patients' understanding of risk associated with medication use: impact of European Commission guidelines and other risk scales. *Drug Saf*. 2003;26(1):1-11.
- Brookes ST, Whitely E, Egger M, Smith GD, Mulheran PA, Peters TJ. Subgroup analyses in randomized trials: risks of subgroup-specific analyses; power and sample size for the interaction test. *J Clin Epidemiol*. 2004 Mar;57(3):229-36.
- Carriero A, de Hond A, Cappers B, Paulovich F, Abeln S, Moons KG, van Smeden M. Explainable AI in healthcare: to explain, to predict, or to describe? *Diagn Progn Res*. 2025 Dec 5;9(1):29.
- Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J Clin Epidemiol*. 2019 Jun;110:12-22.
- Faes L, Sim DA, van Smeden M, Held U, Bossuyt PM, Bachmann LM. Artificial Intelligence and Statistics: Just the Old Wine in New Wineskins? *Front Digit Health*. 2022 Jan 26;4:833912.
- Fingerhut A, Uranues S, Dziri C, Ma J, Vernerey D, Kurihara H, Stiegler P. Interaction analysis of subgroup effects in randomized trials: the essential methodological points. *Sci Rep*. 2024 Jun 1;14(1):12619.
- Fox MP, Adrien N, van Smeden M, Suarez E. Invited commentary: it's not all about residual confounding-a plea for quantitative bias analysis for epidemiologic researchers and educators. *Am J Epidemiol*. 2024 Nov 4;193(11):1609-1611.
- Fu, Edouard and Harhay, Michael and Schneeweiss, Sebastian and Desai, Rishi and Hernán, Miguel A., Starting right: aligning eligibility and treatment assignment at time zero when emulating a target trial (September 13, 2025). Available at SSRN: <https://ssrn.com/abstract=5177135> or <http://dx.doi.org/10.2139/ssrn.5177135>
- Gigerenzer G, Gaissmaier W, Kurz-Milcke E, Schwartz LM, Woloshin S. Helping Doctors and Patients Make Sense of Health Statistics. *Psychol Sci Public Interest*. 2007 Nov;8(2):53-96.
- Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology*. 1999;10(1):37-48.
- Harrell FE Jr, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. *JAMA*. 1982 May 14;247(18):2543-6.
- Hernán, M. A., & Robins, J. M. (2020). *Causal Inference: What If*. Chapman & Hall/CRC
- ICHOM - International Consortium for Health Outcomes Measurement. <https://www.ichom.org/standard-sets/>
- Imbens, G. W., & Rubin, D. B. (2015). *Causal inference for statistics, social, and biomedical sciences: An introduction*. Cambridge University Press.
- Kent DM, Steyerberg E, van Klaveren D. Personalized evidence based medicine: predictive approaches to heterogeneous treatment effects. *BMJ*. 2018 Dec 10;363:k4245.
- Kent DM, Paulus JK, van Klaveren D, D'Agostino R, Goodman S, Hayward R, Ioannidis JPA, Patrick-Lake B, Morton S, Pencina M, Raman G, Ross JS, Selker HP, Varadhan R, Vickers A, Wong JB, Steyerberg EW. The Predictive Approaches to Treatment effect Heterogeneity (PATH) Statement. *Ann Intern Med*. 2020 Jan 7;172(1):35-45.
- Kyriacou DN, Lewis RJ. Confounding by Indication in Clinical Research. *JAMA*. 2016 Nov 1;316(17):1818-1819.
- Little RJ, Rubin DB. Causal effects in clinical and epidemiological studies via potential outcomes: concepts and analytical approaches. *Annu Rev Public Health*. 2000;21:121-45.
- Moons K G M, Damen J A A, Kaul T, Hooft L, Andaur Navarro C, Dhiman P et al. PROBAST+AI: an updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods *BMJ* 2025; 388 :e082505
- Mühlbauer V, Mühlhauser I. Understanding adverse drug reactions in package leaflets – an exploratory survey among health care professionals. *BMC Health Serv Res* 2015;15:505.

- Ramspek CL, Steyerberg EW, Riley RD, Rosendaal FR, Dekkers OM, Dekker FW, van Diepen M. Prediction or causality? A scoping review of their conflation within current observational research. *Eur J Epidemiol*. 2021 Sep;36(9):889-898.
- Riley RD, Snell KIE, Archer L, Ensor J, Debray TPA, van Calster B et al. Evaluation of clinical prediction models (part 3): calculating the sample size required for an external validation study *BMJ* 2024; 384 :e074821
- Riley, RD, Ensor J, Snell KIE, Archer L, Whittle R, Dhiman P. Importance of sample size on the quality and utility of AI-based prediction models for healthcare. *The Lancet Digital Health*, Volume 7, Issue 6, 100857
- Rolf H H Groenwold, Jelle J Goeman, Saskia Le Cessie, Olaf M Dekkers, Multiple testing: when is many too much?, *European Journal of Endocrinology*, Volume 184, Issue 3, Mar 2021, Pages E11–E14
- Stacey, D., Volk, R. J., Smith, M., Lewis, K. B., & Carley, M. for the IPDAS Steering Committee. (2025). Policy Brief: IPDAS 5.0. <https://decisionaid.ohri.ca/ipdas>
- Steyerberg EW. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. 2de dr. Springer Cham; 2019
- Steyerberg EW, Harrell FE Jr. Prediction models need appropriate internal, internal-external, and external validation. *J Clin Epidemiol*. 2016 Jan;69:245-7.
- Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J*. 2014 Aug 1;35(29):1925-31.
- Van Calster, B., McLernon, D.J., van Smeden, M. et al. Calibration: the Achilles heel of predictive analytics. *BMC Med* 17, 230 (2019).
- Van Calster B, Steyerberg EW, Wynants L, van Smeden M. There is no such thing as a validated prediction model. *BMC Med*. 2023 Feb 24;21(1):70.
- van Geloven N, Keogh RH, van Amsterdam W, Cinà G, Krijthe JH, Peek N, Luijken K, Magliacane S, Morzywołek P, van Ommen T, Putter H, Sperrin M, Wang J, Weir DL, Didelez V. The Risks of Risk Assessment: Causal Blind Spots When Using Prediction Models for Treatment Decisions. *Ann Intern Med*. 2025 Sep;178(9):1326-1333.
- van Geloven N, Swanson SA, Ramspek CL, Luijken K, van Diepen M, Morris TP, Groenwold RHH, van Houwelingen HC, Putter H, le Cessie S. Prediction meets causal inference: the role of treatment in clinical prediction models. *Eur J Epidemiol*. 2020 Jul;35(7):619-630.
- Varadhan R, Seeger JD. Estimation and reporting of heterogeneity of treatment effects. In: Velentgas P, Dreyer NA, Nourjah P, et al., eds. *Developing a protocol for observational comparative effectiveness research: a user's guide*. Agency for Healthcare Research and Quality, 2013.
- Veiligheid en toezicht medische hulpmiddelen. Rijksoverheid. <https://www.rijksoverheid.nl/onderwerpen/medische-hulpmiddelen/veiligheid-en-toezicht-medische-hulpmiddelen>
- Veldwijk J, Lambooi MS, van Til JA, Groothuis-Oudshoorn CG, Smit HA, de Wit GA. Words or graphics to present a Discrete Choice Experiment: Does it matter? *Patient Educ Couns*. 2015 Nov;98(11):1376-84.
- Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making*. 2006;26(6):565-574.