Intermediaire uitkomstmaat: verdedigbare keuze?

Tom Poelman, Vakgroep Volksgezondheid en Eerstelijnszorg, UGent

Inleiding

Als alternatief voor een klinisch relevante uitkomstmaat maken onderzoekers vaak gebruik van intermediaire uitkomstmaten die de positieve en/of negatieve effecten van een interventie zouden moeten voorspellen. Het kan hierbij gaan om een causale factor (zoals LDL-cholesterol als surrogaat voor cardiovasculaire uitkomstmaten), een gerelateerde factor (zoals microalbuminurie als surrogaat voor terminaal nierfalen) of een subklinische indicator (zoals coronaire calciumscore (CAC-score) als maat voor vroege vaatveranderingen en als surrogaat voor cardiovasculaire uitkomstmaten). Een gunstig effect van een behandeling op een intermediaire uitkomstmaat betekent echter niet noodzakelijk ook een gunstig effect op de gerelateerde klinische uitkomstmaat (1). Uit een grote meta-epidemiologische studie bleek dat de kans op de rapportering van een positief therapeutisch effect tweemaal groter was in studies met intermediaire uitkomstmaten dan in studies met patiëntrelevante uitkomstmaten (2). Omdat de belangrijkste doelstelling van Minerva erin bestaat om de praktijk van (para-)medici te ondersteunen met klinisch relevante informatie, selecteert de redactie voor duiding in de eerste plaats wetenschappelijke studies met patiëntrelevante uitkomstmaten.

Probleemstelling

In de onderzoekswereld bestaat al langer een spanningsveld tussen enerzijds de voordelen van intermediaire uitkomstmaten, zoals tijdswinst en efficiënter gebruik van middelen, en anderzijds de onzekerheid of het waargenomen effect van de interventie op de intermediaire uitkomstmaat zich ook zal vertalen in een klinisch relevante uitkomst (3-7).

Om de sterkte van het verband tussen een intermediaire uitkomstmaat en een patiëntrelevante uitkomstmaat te kwantificeren bestaan er meerdere statistische modellen*. Momenteel is er echter geen consensus over de meest geschikte statistische methode om intermediaire uitkomstmaten te valideren en evenmin over een drempelwaarde die overschreden moet worden om te besluiten dat de intermediaire uitkomstmaat de klinische uitkomstmaat voldoende betrouwbaar voorspelt.

Voorbeeld:
Volgens de statistische methode van ***surrogate threshold effect (STE)*** *moet de LDL-concentratie met minstens 35 mg/dl dalen om een consistente vermindering van cardiovasculaire gebeurtenissen te verwachten.*

Het beoordelen van de adequaatheid van een intermediaire uitkomstmaat beperkt zich niet alleen tot een statistisch probleem. Zelfs wanneer een intermediaire uitkomstmaat sterk zou correleren met een klinische uitkomstmaat, kan de biologische, klinische en epidemiologische context waarin de intermediaire uitkomstmaat gebruikt wordt de kwantiteit en de kwaliteit van het verband beïnvloeden. Over de mate waarin deze contextfactoren de geloofwaardigheid van intermediaire uitkomstmaten beïnvloeden bestaat er evenmin consensus.

Voorbeelden:
Studies tonen aan dat CRP geassocieerd is met cardiovasculair risico (8), maar het is minder duidelijk of inflammatie een biologische oorzaak is van cardiovasculaire ziekte. Dat maakt CRP als intermediaire uitkomstmaat minder geloofwaardig dan uitkomstmaten waarvoor een sterkere biologische en klinische validatie bestaat.
Terminaal nierfalen is relatief zeldzaam en ontwikkelt zich vaak langzaam (9). Daarom zijn er grote studies met een lange follow-up nodig om het verband tussen een intermediaire uitkomstmaat en terminaal nierfalen aan te tonen. Deze epidemiologische context bemoeilijkt de validatie van een intermediaire uitkomstmaat voor terminaal nierfalen..

Ontwikkeling van een raamwerk voor evaluatie van intermediaire uitkomstmaten

In een poging om tot een consensus te komen over de evaluatie van intermediaire uitkomstmaten werd het *REASSurE (Recommendations for Evaluating A Single Surrogate Endpoint)* project opgezet (10).

Aansluitend bij de geschetste probleemstelling hierboven formuleerden de onderzoekers drie kernvragen bij de evaluatie van intermediaire uitkomstmaten:

1. Wat zijn de belangrijkste niet-statistische factoren?
2. Moeten er numerieke drempels gebruikt worden om de validiteit van een intermediaire uitkomstmaat te beoordelen?
3. Met welke statistische methode kan men het best een intermediaire uitkomstmaat valideren?

Op basis van deze kernvragen stelden ze een vragenlijst op die in twee Delphi-ronden gevalideerd werd door 5 groepen van stakeholders: statistici, onderzoeksmethodologen, regelgevers, HTA (Health Technology Assessment)-experten en clinici. Na een internationale online consensus meeting werd er overeenstemming bekomen over een raamwerk voor de evaluatie van intermediaire uitkomstmaten.

Het REASSurE (Recommendations for Evaluating A Single Surrogate Endpoint)-framework als resultaat

Domein 0: Beschrijft de klinische context waarin de intermediaire uitkomstmaat zal gebruikt worden:

- Wat zijn de kenmerken van de onderzochte behandeling (indicatie, geneesmiddelklasse, behandellijn)?
- Wat zijn de gebruikte intermediaire en klinische uitkomstmaten in de studie?
- In welke fase (1,2,3,4) van het klinisch onderzoek zal men de intermediaire uitkomstmaat gebruiken?

*Een correcte beschrijving hiervan is belangrijk om te kunnen inschatten of een intermediaire uitkomstmaat voor een bepaalde studie überhaupt noodzakelijk is. Zo is het gebruik van een intermediaire uitkomstmaat meer te verantwoorden in een **fase 2-studie** omdat men hier vooral wil nagaan of een behandeling biologisch werkzaam is. Ook voor een tweedelijnsbehandeling zal een intermediaire uitkomstmaat sneller aanvaard worden omdat de groep patiënten waarvoor de eerstelijnsbehandeling niet effectief is, te klein is om met voldoende power een effect op een harde, klinische uitkomstmaat aan te tonen.*

Domein 1: Praktische motivatie

- Item 1.1: Ziekteprevalentie: zeldzaam versus veelvoorkomend?

Aanbevolen item om te evalueren; Bij zeldzame aandoeningen kan een intermediaire uitkomstmaat sneller verantwoord zijn omdat het vaak niet mogelijk is om voldoende patiënten te rekruteren en lang genoeg op te volgen om met voldoende statistische power een effect op harde klinische eindpunten aan te tonen.

- Item 1.2: Klinische noodzaak: dringend of onvervuld?

Aanbevolen item om te evalueren; Bij een ernstige aandoening met grote ziektelast voor de patiënt waarvoor er nog geen doeltreffende behandeling voorhanden is, kan een intermediaire uitkomstmaat meer aanvaard zijn wanneer de verwachte voordelen van een snellere toegang tot de behandeling opwegen tegenover de onzekerheid over het effect op harde klinische eindpunten.

- Item 1.3: Tijdswinst: wel of niet de moeite waard?; onbepaald?

***Verplicht** item om te evalueren; Een intermediaire uitkomstmaat kan de duur van een klinische studie verkorten omdat het effect ervan sneller meetbaar is. De tijdswinst moet echter steeds afgewogen worden tegen de statistische onzekerheid (zie domein 3) met betrekking tot het werkelijke effect van de interventie op harde klinische eindpunten.*

- Item 1.4: Andere uitgespaarde middelen: wel of niet de moeite waard?; onbepaald?

Aanbevolen item om te evalueren; Het gebruik van een intermediaire uitkomstmaat kan gezondheidswinst voor patiënten opleveren en de onderzoeks- en zorgkosten doen dalen. De kosteneffectiviteit moet echter worden

geëvalueerd en afgewogen tegen de statistische onzekerheid (zie domein 3) over het werkelijke effect van de interventie op harde klinische eindpunten.

Domein 2: Klinisch en biologisch bewijs over een verband tussen de intermediaire uitkomstmaat en een patiëntrelevante uitkomstmaat

- Item 2.1: Bewijs uit meta-analyse(s): hoeveel?, methodologische kwaliteit, aantal geïncludeerde studies en patiënten, rapportering van de sterkte van het verband tussen intermediaire uitkomstmaat en patiëntrelevante uitkomstmaat, zowel op patiëntniveau (hebben patiënten met een gunstige verandering in de intermediaire uitkomstmaat ook een betere klinische uitkomst?) als op studieniveau (hebben studies waarin de behandeling een groter effect heeft op de intermediaire uitkomstmaat ook een groter effect op de klinische uitkomst?); alsook consistentie van de resultaten over verschillende meta-analyses heen?

- Item 2.2: Bewijs uit RCT(s): hoeveel? (dat nog niet in een meta-analyse is opgenomen), methodologische kwaliteit, steekproefgrootte, rapportering van de sterkte van het verband tussen intermediaire uitkomstmaat en klinisch relevante uitkomstmaat, zowel op patiëntniveau als op studieniveau; alsook consistentie van de resultaten over verschillende RCT's heen?

*Om een intermediaire uitkomstmaat te mogen gebruiken **moet** haar validiteit als surrogaat voor een patiëntrelevante uitkomst minstens in een meta-analyse en/of een RCT zijn aangetoond.*
Bijvoorbeeld: Om te weten of men LDL-cholesterol als primaire uitkomstmaat mag gebruiken om het effect van een nieuw geneesmiddel te testen, moet men eerst nagaan of er voldoende bewijs bestaat dat andere behandelingen die LDL-cholesterol verlaagden ook het aantal cardiovasculaire aandoeningen deden dalen.

- Item 2.3: Observationele studies

*Een verband tussen een intermediaire uitkomstmaat en een patiëntrelevante uitkomstmaat dat alleen onderbouwd is met een observationele studie **kan nuttig zijn maar is onvoldoende als bewijs.***

- Item 2.4: Biologische plausibiliteit

***Verplicht** item; Is er een biologische verklaring (zoals pathogenese van de aandoening) voor het verband tussen een verandering in de intermediaire uitkomstmaat en een verandering in een patiëntrelevante uitkomstmaat?*

- Item 2.5: Klinische plausibiliteit

***Verplicht** item; Maakt de intermediaire uitkomstmaat deel uit van het causale ziekteproces en vertaalt een verandering in deze uitkomstmaat zich ook in een verandering van de klinische uitkomst?*

- Item 2.6: Generaliseerbaarheid

***Verplicht** item; Is er bewijs dat het gebruik van de intermediaire uitkomstmaat ook naar andere behandelklassen en/of populaties geëxtrapoleerd kan worden?*

Domein 3: Statistische methodes

- Item 3.1: Analyse op individueel niveau

Deze analyse onderzoekt het verband tussen de intermediaire uitkomstmaat en de klinische uitkomst op individueel niveau. Een intermediaire uitkomstmaat met een sterke prognostische waarde bewijst echter niet noodzakelijk dat een positief effect van de behandeling ook de klinische uitkomst zal verbeteren. Bovendien zijn individuele patiëntgegevens niet altijd beschikbaar. Daarom wordt deze analysetechniek door de consensusgroep niet naar voor geschoven.

- Item 3.2: Analyse op studieniveau

Deze analyse onderzoekt of behandelingen die een groter effect hebben op de intermediaire uitkomstmaat ook consequent leiden tot een groter effect op de klinische uitkomst. Omdat hiermee kan worden nagegaan in welke mate het behandelingseffect op de klinische uitkomstmaat kan worden voorspeld door het behandelingseffect op de intermediaire uitkomstmaat, beschouwen de stakeholders van het project deze analysemethode als onontbeerlijk voor de evaluatie van een intermediaire uitkomstmaat. Bij analyse op studieniveau moet dan ook steeds de R^2-waarde vermeld worden. R^2 geeft aan welk aandeel van de variatie tussen studies in het behandelingseffect op de klinische uitkomst wordt verklaard door het behandelingseffect op de intermediaire uitkomstmaat. Deze maat beschrijft dus de sterkte van de associatie tussen beide effecten op studieniveau. $R^2=1$ betekent hierbij een perfecte voorspelling van het behandelingseffect op de klinische uitkomst, terwijl $R^2=0$ aangeeft dat de intermediaire uitkomstmaat geen voorspellende waarde heeft.

- Item 3.3: R^2-drempelwaarden (R^2 thresholding)

Er bestaan geen algemeen aanvaarde drempelwaarden voor R^2 en zelfs wanneer men een bepaalde R^2 als voldoende hoog beschouwt, moet men rekening blijven houden met andere (contextuele) factoren.

Besluit

Een internationale groep van stakeholders, waaronder ook clinici, kwam via een Delphi-procedure en een consensusvergadering tot een raamwerk voor de evaluatie van intermediaire uitkomstmaten. Een intermediaire uitkomstmaat is alleen aanvaardbaar wanneer de bekomen tijdswinst opweegt tegen de toegenomen onzekerheid over het effect op patiëntrelevante uitkomstmaten. Het verband tussen een intermediaire uitkomstmaat en een patiëntrelevante uitkomstmaat moet biologisch en klinisch plausibel zijn, alsook voldoende statistisch onderbouwd in RCT's en meta-analyses. Het is aanbevolen om daarnaast ook de prevalentie van de aandoening, de klinische noodzaak, de afweging tussen bespaarde middelen en onzekerheid, alsook beschikbaar bewijs uit observationele studies mee in overweging te nemen. Er bestaat geen consensus over de keuze van de statistische methode om de validiteit van een intermediaire uitkomstmaat te evalueren. Een analyse op studieniveau wordt wel als essentieel beschouwd, terwijl een analyse op individueel niveau aanvullende informatie kan opleveren. Drempelwaarden voor de correlatie tussen het behandelingseffect op de intermediaire uitkomstmaat en het behandelingseffect op de patiëntrelevante uitkomstmaat kunnen nuttig zijn, maar mag men nooit loskoppelen van de contextfactoren.

**zoals: Proportion of Treatment Effect Explained (PTE) (Welk percentage van het effect op de klinische uitkomstmaat kan verklaard worden door het effect op de intermediaire uitkomstmaat?); Individual-level association (Leidt een betere score op een intermediaire uitkomstmaat ook tot een betere klinische uitkomst bij een patiënt?); Trial-level association (Hebben studies waarin de behandeling een groot effect heeft op de intermediaire uitkomstmaat ook een groot effect op de klinische uitkomst?); Surrogate Threshold Effect (STE) (Hoe groot moet het effect op de intermediaire uitkomstmaat zijn vooraleer we met voldoende zekerheid een gunstig effect op de klinische uitkomst mogen verwachten?); Causal effect predictiviness (Voorspelt een verandering in de intermediaire uitkomstmaat door de behandeling ook een verandering in klinische uitkomst bij een patiënt?); Individual causal association (Is de causale invloed van de behandeling op de intermediaire uitkomstmaat gekoppeld aan de causale invloed op de klinische uitkomst bij dezelfde patiënt?); Sample Size Multiplier (Hoeveel meer patiënten zijn er nodig als we de klinische uitkomstmaat gebruiken in plaats van de intermediaire uitkomstmaat?); True Endpoint Advantage (Wat winnen we door rechtstreeks de klinische uitkomst te meten in plaats van de intermediaire uitkomstmaat?)*

Referenties

1. Chevalier P. Intermediaire uitkomstmaten: beperkte betrouwbaarheid van de resultaten. Minerva 2014;13(7):90.
2. Chevalier P. Intermediaire uitkomstmaten versus klinisch relevante uitkomstmaten: verschil in effectgrootte? Minerva 2014;13(9):116.
3. Boissel JP, Collet JP, Moleur P, Haugh M. Surrogate endpoints: a basis for a rational approach. Eur J Clin Pharmacol 1992;43:235-44. DOI: 10.1007/BF02333016

4. Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled? Ann Intern Med 1996;125:605-13. DOI: 10.7326/0003-4819-125-7-199610010-00011
5. Alonso A, Molenberghs G. Surrogate end points: hopes and perils. Expert Rev Pharmacoecon Outcomes Res 2008;8:255-9. DOI: 10.1586/14737167.8.3.25
6. Patel RB, Vaduganathan M, Samman-Tahhan A, et al. Trends in utilization of surrogate endpoints in contemporary cardiovascular clinical Trials. Am J Cardiol 2016;117:1845-50. DOI: 10.1016/j.amjcard.2016.03.021
7. Manyara AM, Davies P, Stewart D, et al. Definitions, acceptability, limitations, and guidance in the use and reporting of surrogate end points in trials: a scoping review. J Clin Epidemiol 2023;160:83-99. DOI: 10.1016/j.jclinepi.2023.06.013
8. Yang X, Zhang D, Zhao Y, et al. Association between serum level of C-reactive protein and risk of cardiovascular events based on cohort studies. J Hum Hypertens 2021;35:1149-58. DOI: 10.1038/s41371-021-00546-z
9. Liu P, Quinn RR, Lam NN, et al. Progression and regression of chronic kidney disease by age among adults in a population-based cohort in Alberta, Canada. JAMA Netw Open 2021 Jun;4:e2112828. DOI: 10.1001/jamanetworkopen.2021.12828
10. Lo WP, Baker SG, Bertolaccini L, et al. Recommendations for evaluating a single surrogate endpoint (REASSurE): consensus based, surrogate evaluation framework with explanation and elaboration. BMJ 2026;394:e100618. DOI: 10.1136/bmj-2026-100618