

Machine learning vereist betere risicoverevening zorgverzekeraars

Zorgverzekeraars worden gecompenseerd voor de risicoprofielen van hun verzekerden. Het huidige risicovereveningsmodel corrigeert echter niet voor alle gezondheidsverschillen. Met behulp van machine learning-technieken (ML) zouden zorgkosten beter voorspeld kunnen worden. Zou ML gebruikt kunnen worden voor risicoselectie, en wat kan de overheid hier tegen doen?

IN HET KORT

- ML-technieken kunnen zorgkosten beter voorspellen dan de gebruikelijke lineaire regressietechnieken.
- Met de betere voorspellingen kunnen zorgverzekeraars gunstige risico's selecteren, wat maatschappelijk ongewenst is.
- Om risicoselectie tegen te gaan dient de overheid de lessen van ML toe te passen bij de risicoverevening tussen verzekeraars.

PIET STAM

Partner bij Equalis en gastonderzoeker aan de Vrije Universiteit Amsterdam (VU)

ISMAIL

Onderzoeker aan de VU

JOHAN VISSER

Senior consultant bij Equalis

FRANCE PORTRAIT

Universitair hoofd-docent aan de VU

XANDER KOOLMAN

Universitair hoofd-docent aan de VU

Van zorgverzekeraars wordt er sinds de invoering van de Zorgverzekeringswet (Zvw) in 2006 verwacht dat zij, voor hun verzekerden, betaalbare, toegankelijke en kwalitatief goede zorg organiseren. Jaarlijks strijden de zorgverzekeraars op de zorgverzekeringsmarkt om de gunst van de verzekerden door hun een aantrekkelijk aanbod te doen tegen een betaalbare premie. De Zvw heeft wel een aantal beperkingen opgelegd aan de concurrentie tussen zorgverzekeraars (Zorgverzekeringswet, 2004; Stam, 2015). Door de acceptatieplicht en het algehele verbod op premiedifferentiatie ontstaan er voorspelbare winsten en verliezen binnen een polis. Hierdoor ontstaat er voor de verzekeraars een financiële prikkel om zich meer te richten op verzekerden met een gunstig risicoprofiel dan op die met een ongunstig risicoprofiel. Zorgverzekeraars kunnen bijvoorbeeld door directe marketing inspelen op deze prikkel tot risicoselectie (NZa, 2019).

Om deze financiële prikkel tegen te gaan, is er landelijk een model van risicoverevening ingevoerd. Door de werking van dit model ontvangen zorgverzekeraars naast de nominale premies ook compensaties uit het Zorgverzekeringsfonds, om ervoor te zorgen dat alle verzekerden financieel even aantrekkelijk zijn. Voor verzekerden met een ongunstig risicoprofiel ontvangen verzekeraars een hogere compensatie dan voor hen met een gunstig risicoprofiel.

Over- en ondercompensaties

Het kan zijn dat er ondanks de geavanceerde vereveningsformule sprake is van over- of ondercompensatie van verzekerden. In dat geval blijft een deel van de prikkel tot risicoselectie bestaan. Dit is een probleem voor zover de risicoverevening daarvoor behoort te compenseren (Stam et al., 2015), en verzekeraars niet in staat zijn om de over- en ondercompensaties weg te nemen via beleidsaanpassingen (bijvoorbeeld doelmatige zorginkoop).

Als zorgverzekeraars over de instrumenten beschikken om deze groepen vast te stellen, kunnen ze de financiële prikkel tot risicoselectie daadwerkelijk ook in acties omzetten (Visser et al., 2014). Er is al langer bekend dat zorgverzekeraars substantieel en structureel worden ondercompenseerd voor chronisch zieken, hoe geavanceerd het risicovereveningsmodel tot nu toe ook is (Stam en Van de Ven, 2006; 2007; 2008; Van Kleef et al., 2012; 2014).

Met de afbouw van de reserves van zorgverzekeraars in de afgelopen jaren hebben zorgverzekeraars minder mogelijkheden om een negatief resultaat op chronisch zieken te compenseren. Hierdoor lijken de premieverschillen tussen zorgverzekeraars met verschillende verzekerdenpopulaties, mede door deze ondercompensatie van chronisch zieken, steeds meer toe te nemen (Kiers, 2019; Bruins, 2019).

Deze toename in premieverschillen suggereert een urgentie om de compensatie voor chronisch zieken in de risicoverevening op korte termijn te verbeteren. Het perspectief op basis van bestaande econometrische modellen is de afgelopen jaren echter beperkt gebleken, waardoor er behoefte is aan een nieuwe oplossingsrichting.

Door gebruik te maken van ML-technieken kan achterhaald worden bij welke groepen verzekerden er (structureel) sprake is van over- of ondercompensatie. Deze informatie kunnen verzekeraars gebruiken om risicoselectie toe te passen. De overheid kan deze informatie gebruiken om (structureel) over- en ondercompensatie te voorkomen.

Het huidige risicovereveningsmodel

De huidige vereveningsformule is gebaseerd op gegevens van alle Nederlandse inwoners over hun leeftijd, geslacht en gezondheidskenmerken. Deze gegevens bestaan uit de declaraties van de zorgverzekeraars, de Belastingdienst, het Uitvoeringsinstituut Werknemersverzekeringen (UWV) en Dienst Uitvoering Onderwijs (DUO).

Zorginstituut Nederland stelt de vereveningsbijdrage



van de zorgverzekeraars vast op basis van de risicokenmerken 'leeftijd', 'geslacht' en 'gezondheid'. Met behulp van regressietechnieken wordt er uitgerekend wat de verwachte zorgkosten zijn voor iedere individuele verzekerde op basis van deze kenmerken. Deze verwachte zorgkosten vormen de basis voor de compensaties richting de zorgverzekeraars. Sinds 2006 worden deze kenmerken jaarlijks onderhouden. Zo worden bijvoorbeeld kenmerken verwijderd die niet langer relevant zijn, en zijn er nieuwe kenmerken toegevoegd die de kosten beter voorspellen.

Risicoverevening met machine learning

Als alternatief voor het gebruik van de reguliere lineaire regressietechnieken (OLS) zijn we nagegaan of de zorgkosten beter voorspeld kunnen worden als we ML-algoritmen gebruiken. ML kan meerwaarde opleveren omdat de relevante interactie-effecten tussen de kenmerken in het risicovereveningsmodel geautomatiseerd worden bepaald.

In ons onderzoek maken we gebruik van de landelijke risicovereveningsdata van het model van 2018, met dank aan het Ministerie van VWS en Zorgverzekeraars Nederland. Wij zijn uitgegaan van de bestaande verzameling kenmerken in dat model. De meeste interacties zijn niet eerder onderzocht.

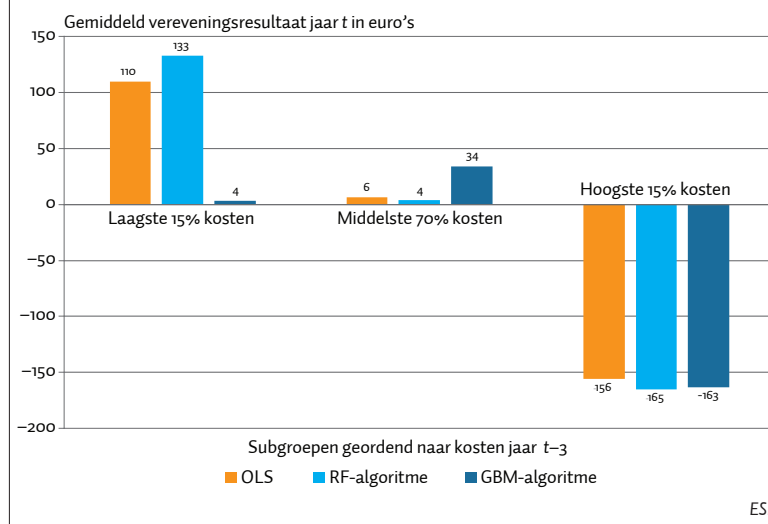
Twee ML-algoritmen die een rol kunnen spelen bij een verbetering van het signaleren van eventuele relevante interacties zijn Random Forests (RF's) en Gradient Boosted Machines (GBM's) (Ismail, 2018). Deze technieken zijn een logisch vervolg op de eerdere pogingen met regressiebomen (*regression trees*) in de risicoverevening (Van Veen, 2017). Een regressieboom kan gezien worden als een verzameling 'if-then'-statements, waarmee data stapsgewijs opgedeeld worden in steeds kleinere homogene groepen. Enkelvoudige regressiebomen hebben echter de neiging om zich te sterk aan te passen aan de rol van het toeval in de data waarop ze getraind zijn ('overfitting'), waardoor ze slechter presteren op nieuwe data. De RF- en GBM-algoritmes combineren meerdere regressiebomen, die onderling van elkaar afwijken, tot zogeheten 'ensembles'. De verschillen tussen de regressiebomen ontstaan bij RF op een andere wijze dan bij GBM. De toevalsvariatie tendeert naar nul, naarmate het aantal gecombineerde regressiebomen verder toeneemt. Daardoor leveren RF en GBM voorspel-

lingen op die beter generaliseerbaar zijn naar nieuwe data dan enkelvoudige regressiebomen.

We hebben RF's en GBM's toegepast om de somatische zorgkosten – de zorgkosten van chronisch lichamelijke aandoeningen – van de Nederlandse populatie in 2018 te voorspellen. Deze voorspellingen hebben we op individueel- (zie methode in kader 1) en subgroepniveau (zie methode in kader 2) vergeleken met de voorspellingen die resulteren als de gebruikelijke lineaire regressiemethode (OLS) wordt toegepast.

Vereveningsresultaten van subgroepen met OLS, RF en GBM

FIGUUR 1



Evaluatie van voorspelkracht van OLS, RF en GBM op individueel niveau

TABEL 1

Evaluatiemaat ¹	OLS	RF-algoritme	GBM-algoritme
R-kwadraat (procent)	32,0	33,5	32,4
CPM (procent)	31,9	32,7	30,0
GGAA (€)	1.984	1.961	2.029

¹ De evaluatiematen worden verder toegelicht in kader 1.

Voorspelkracht van de verschillende modellen

Figuur 1 toont de vereveningsresultaten van 2018 voor de subgroepen met de laagste en hoogste vijftien procent, en de middelste zeventig procent kosten drie jaar daarvoor. Hoe dichter het vereveningsresultaat bij de 0 is, hoe minder er gemiddeld per individu in deze subgroepen sprake is van over- of ondercompensatie – en dus ook van een prikkel voor risicoselectie.

Voorspellingen van machine learning op individueel niveau

KADER 1

De statistische evaluatie van de voorspelkracht van de modellen op individueel verzekerniveau is uitgerekend aan de hand van de bij de risicoverevening gebruikelijke evaluatiematen R-kwadraat, de voorspelmaat van Cumming (*Cumming's prediction measure*; CPM) en de gewogen gemiddelde absolute voorspelfout (GGAA).

De R-kwadraat geeft aan welk deel van de variatie in zorgkosten door onze modellen wordt voorspeld. Dat geldt ook voor de CPM, maar in dat geval wordt de variatie in de absolute in plaats van de gekwadrateerde verschillen tussen zorgkosten en modelvoorspellingen uitgedrukt. Beide maten variëren tussen de 0 procent en de 100 procent: hoe hoger de score, des te preciezer de voorspellingen aansluiten bij de variatie in de zorgkosten. De GGAA geeft aan hoe groot de absolute verschillen tussen de zorgkosten en modelvoorspel-

lingen zijn, en wordt uitgedrukt in euro's. Hierbij geldt dat hoe kleiner de GGAA, hoe beter de genoemde aansluiting is. Deze voorspelmaten zijn op individueel verzekerniveau uitgerekend.

Om te voorkomen dat het ML-algoritme zich te veel aanpast aan de kostenpatronen in onze totale dataset van zeventien miljoen individuele verzekerden ('overfitting'), hebben we de modellen geschat op een random steekproef (zeventig procent) van deze data, en zijn zo de evaluatiematen op basis van het restant (dertig procent) berekend. Als we dit allebei op hetzelfde totale bestand hadden gedaan, dan zouden de evaluatiematen naar verwachting hoger zijn uitgevallen, maar dan zouden we geen antwoord hebben gekregen op de vraag hoe generaliseerbaar deze modeluitkomsten naar andere, recentere datasets zijn.

Voorspellingen van machine learning op subgroepniveau

KADER 2

Voor de evaluatie van een risicovereveningsmodel op subgroepniveau, wordt de kwaliteit beoordeeld van de voorspelling van zorgkosten voor geselecteerde subgroepen van verzekerden. Deze geselecteerde subgroepen worden samengesteld op basis van informatie voorafgaand aan het jaar van de realisatie van de zorgkosten, omdat het risicovereveningssysteem geacht wordt compensatie te bieden voor de voorspelbare kosten van verzekerden. Immers, prikkels tot risicoselectie kunnen alleen bestaan ten aanzien van kosten van de geselecteerde subgroepen die voorspelbaar zijn.

We hebben ervoor gekozen om de zorgkosten van de subgroepen te voorspellen die zijn samengesteld op basis van kostengegevens van de drie jaar voorafgaand aan het vereveningsjaar 2018. Hierdoor ontstaat inzicht in de ondergrens van wat er met deze technieken kan worden bereikt. Met recentere gegevens zouden we een nog scherper beeld kunnen krijgen van de mogelijkheden van ML, maar die zijn niet landelijk beschikbaar. Zorgverzekeraars hebben wel recentere gegevens beschikbaar van twee of één jaar voorafgaand aan het vereveningsjaar, waardoor de moge-

lijkheden voor zorgverzekeraars om met RF en GBM groepen van ondergecompenseerde verzekerden vast te stellen naar verwachting groter zullen zijn dan uit onze analyse blijkt.

In de risicoverevening wordt er een ondergrens van vijftig euro meer- of minderkosten per verzekerde gehanteerd bij de statistische beoordeling of een subgroep opgenomen moet worden in het vereveningsmodel. Onder meer om deze reden ontstaan er over- en ondercompensaties van subgroepen die niet in het vereveningsmodel worden opgenomen. Voor zover zorgverzekeraars over- en ondercompensaties doorberekenen in hun premies, kunnen die op de polismarkt aanleiding geven voor verzekerden om over te stappen van de ene naar de andere verzekeraar. In de overstapperiode voor 2018/2019 leidde een stijging van de premie met tien procent – in termen van de gemiddelde premie in 2019 was dat 138 euro – ertoe dat zeven procent van de verzekerden met de desbetreffende polis overstapte (NZA, 2019). Ten aanzien van de omvang van de bedragen in figuur 1 mag er dus worden verondersteld dat deze relevant zijn voor het premiebeleid van zorgverzekeraars.

In het risicovereveningsmodel 2018 berekend met OLS, blijken de subgroepen met de laagste vijftien procent en de middelste zeventig procent kosten van drie jaar daarvoor te worden overgecompenseerd met respectievelijk 110 en 6 euro per verzekerde. Maar de subgroep met de hoogste vijftien procent kosten van drie jaar daarvoor werd juist ondergecompenseerd met 156 euro per verzekerde.

Wanneer RF wordt gebruikt om de risicoverevening te berekenen, blijkt dat de overcompensatie in 2018 voor de brede middengroep de helft kleiner is. Dit gaat ten koste van het resultaat voor de overige twee subgroepen, waarvoor de overcompensatie (laagste vijftien procent kosten) en de ondercompensatie (hoogste vijftien procentkosten) juist toenemen. Uit tabel 1 volgt dat deze herverdeling van de compensaties tussen de subgroepen tot een vergroting van de totale voorspelkracht van het risicovereveningsmodel leidt. Immers, alle evaluatiematen van het RF-algoritme pakken beter uit dan die van OLS en GBM, wat voor toepassing van RF pleit.

Uit figuur 1 blijkt verder dat toepassing van GBM de compensaties van de subgroep met de laagste kosten drie jaar voor verevening sterk verkleint, in plaats van bij de middengroep zoals bij RF. De overcompensatie van 110 euro van deze subgroep mitigeert door GBM tot 4 euro. Hierdoor neemt de prikkel tot selectie af van de relatief gezonde verzekerden, waar nu de meeste marketinginspanningen op zijn gericht. Dit gaat ten koste van het resultaat voor de overige twee subgroepen, waarbij het opvalt dat vooral de overcompensatie van de middengroep (middelste zeventig procent kosten) sterk toeneemt. Uit tabel 1 volgt dat deze herverdeling van de compensaties tussen de subgroepen niet zonder meer de totale voorspelkracht van het risicovereveningsmodel verbetert. Immers, alleen de R-kwadraat is bij het GBM-algoritme beter dan bij OLS.

Een verklaring voor de verschillen tussen de vereveningsresultaten met RF en GBM kan zijn dat het GBM-algoritme mogelijk inherent leidt tot een focus op de delen van de populatie waar de residuen het hoogst zijn, in ons geval dus op de groepen met de vijftien procent hoogste of laagste kosten. De toekenning van een hoger gewicht aan de extremen kan evenzeer een verklaring zijn voor de hogere R-kwadraat bij GBM dan bij OLS, terwijl de CPM en GGAA juist slechter uitpakken bij GBM. Immers, extreme verschillen tussen de compensaties en zorgkosten tellen in termen van de R-kwadraat zwaarder (gekwadrateerde verschillen) mee dan bij de overige twee evaluatiematen (absolute verschillen).

Noodzaak verbetering risicoverevening

Toepassing van ML-algoritmen in de risicoverevening leidt tot betere compensaties voor subgroepen van verzekerden, en daarmee tot een gelijkjer speelveld voor zorgverzekeraars. Tegelijkertijd is met de komst van deze technieken ook de noodzaak tot het verbeteren van de risicoverevening groter geworden. Immers, ook zorgverzekeraars kunnen met behulp van ML de voorspelbare winst- en verliesgevende verzekerden gemakkelijker vaststellen.

Wij hebben laten zien dat zorgverzekeraars de mogelijkheid hebben om met deze ML-technieken het huidige risicovereveningsmodel te verslaan, zelfs zonder de variabe-

len aan te passen of andere dan de huidige kenmerken in het risicovereveningsmodel op te nemen.

De kennis, vaardigheden en tools om nieuwe ML-technieken toe te passen, nemen snel toe. Zo zet Vektis inmiddels kunstmatige-intelligentie-technieken in om zorgverzekeraars in staat te stellen de behoefte aan wijkverpleging beter te voorspellen (Adema, 2020). De informatie die dat oplevert kan eveneens worden ingezet voor een betere zorginkoop, maar vergroot ook de mogelijkheden van risicoselectie, omdat het duidelijker wordt bij welke groepen verzekerden er sprake is van ondercompensatie.

Het is de verantwoordelijkheid van de overheid om prikkels tot risicoselectie zo veel mogelijk weg te nemen, bijvoorbeeld door het systeem van risicoverevening verder te optimaliseren. Het gebruik van ML-technieken kan bijdragen aan die verdere optimalisering.

Naar toepassing van machine learning

Vooralsnog zijn de uitkomsten van ML voor velen nog een *black box*, wat voor de overheid een reden is om bij de uitvoering van de risicoverevening vast te houden aan de lineaire regressietechnieken (OLS). Het is voor de overheid immers van belang om transparant te zijn over de gehanteerde methoden. Ook moeten de resultaten voor verzekeraars voorspelbaar zijn, zodat ze in staat worden gesteld hierop beleid te maken. Dat betekent dat het sec toepassen van ML-technieken (RF en GBM) binnen de risicoverevening op basis van de huidige wet- en regelgeving niet goed mogelijk is voor de overheid.

Het is hoopvol dat er recentelijk veelbelovende methoden zijn ontwikkeld om de interacties uit de 'black box' van het ML-algoritme expliciet te maken, zodat de inzichten uit ML-technieken in de huidige regressietechnieken voor risicoverevening kunnen worden toegepast. Equalis en de Vrije Universiteit Amsterdam voeren op dit moment onderzoek uit naar het toevoegen van zulke interacties aan de huidige kenmerken van het risicovereveningsmodel, waarbij de gebruikelijke methode (OLS) gehandhaafd blijft.

De voorlopige resultaten stemmen ons hoopvol dat de compensaties die uit deze aanpak volgen een goede benadering vormen van de compensaties die het resultaat zijn van een directe toepassing van ML-algoritmen. Zo kan men de potentie van het gebruik van ML voor risicoselectie een stap voorblijven, zonder dat het de uitvoering van de risicoverevening complexer en minder uitlegbaar maakt. We roepen de overheid op om te investeren in kennisontwikkeling in de toepassing van ML in de risicoverevening.

Daarnaast behoeft de samenstelling van de kostenvariabelen ook meer onderzoek. In dit onderzoek waren de ondercompensaties van chronisch zieken de directe aanleiding om de compensaties vanuit de risicoverevening met ML te verbeteren. Het is echter niet uitgesloten dat deze ondercompensaties ten minste deels zijn gebaseerd op kostenvariatie waarvoor de risicoverevening uitdrukkelijk niet beoogt te compenseren, zoals zorgaanbodverschillen tussen regio's. De risicoverevening behoort zorgverzekeraars bijvoorbeeld niet te compenseren voor kostenvariatie die wordt veroorzaakt door ondoelmatige praktijkvariatie tussen zorgverleners (Stam et al., 2015). Anders zou de risicoverevening de prikkels tot doelmatigheid teniet doen (Koerhuis et al.,

2016). Wij roepen de overheid op om aanvullend onderzoek te doen naar deze onbedoelde compensaties, zowel ten aanzien van de huidige vereveningskenmerken als bij de gevonden interacties daartussen bij toepassing van ML-algoritmen.

Literatuur

- Adema, J. (2020) *Met AI zorgkosten voorspellen en kwaliteit verbeteren*. Vektis, blog, 15 april.
- Bruins, B. (2019) *Antwoorden op de vragen van het Kamerlid Ellemeet (GL) over het bericht 'Samenstelling verzekerdenpopulatie bepaalt resultaat zorgverzekeraars' (2019Z05385)*. Kamerbrief, 19 maart. Te vinden op www.rijksoverheid.nl.
- Ismail, I. (2018) *Improving risk equalization through machine learning: a comparative evaluation of Random Forests and Gradient Boosted Machines to OLS regression*. Master thesis Health Sciences, VU Amsterdam.
- Kiers, B. (2019) *Samenstelling verzekerdenpopulatie bepaalt resultaat zorgverzekeraars*. Artikel op www.zorgvisie.nl, 14 maart 2019.
- Kleef, R.C. van, R.C.J.A. van Vliet en W.P.M.M. van de Ven (2012) *Risicoverevening 2012: een analyse van voorspelbare winsten en verliezen op subgroepniveau*. Instituut Beleid & Management Gezondheidszorg, 2012.07. Te vinden op www.eur.nl.
- Kleef, R.C. van, R.C.J.A. van Vliet en W.P.M.M. van de Ven (2014) *Risicoverevening 2014 voor somatische zorg: analyse van uitkomsten op subgroepniveau*. Instituut Beleid & Management Gezondheidszorg, 2014.02. Te vinden op www.eur.nl.
- Koerhuis, S., J.S. Visser, X. Koolman en P.J.A. Stam (2016) *Betere risicoverevening kan zorgverzekeraars prikkelen tot meer doelmatigheid*. ESB, 101(4729), 172–175.
- NZa (2019) *Monitor Zorgverzekeringen 2019*. Nederlandse Zorgautoriteit, september. Te vinden op puc.overheid.nl.
- Stam, P.J.A. en W.P.M.M. van de Ven (2006) *Risicoverevening in de zorgverzekering: een evaluatie en oplossingsrichtingen voor verbetering*. Instituut Beleid & Management Gezondheidszorg. Te vinden op www.eur.nl.
- Stam, P.J.A. en W.P.M.M. van de Ven (2007) *Evaluatie risicoverevening: prikkels tot risicoselectie? Een evaluatie van het vereveningsmodel 2007 en oplossingsrichtingen voor verbetering*. Instituut Beleid & Management Gezondheidszorg. Te vinden op www.eur.nl.
- Stam, P.J.A. en W.P.M.M. van de Ven (2008) *De harde kern in de risicoverevening*. ESB, 93(4529), 104–107.
- Stam, P.J.A., J.S. Visser, en R. Goudriaan (2015) *Risicoverevening is geen panacee*. ESB, 100(4720), 624–627.
- Veen, S.H.C.M. van, R.C. van Kleef, W.P.M.M. van de Ven en R.C.J.A. van Vliet (2017) *Exploring the predictive power of interaction terms in a sophisticated risk equalization model using regression trees*. *Health Economics*, 27, 1–12.
- Visser, J.S., J. Sonneveld en P.J.A. Stam (2014) *Het voorkomen van inadequate compensatie in de risicoverevening*. SiRM - Strategies in Regulated Markets, Rapport 16 oktober. Te vinden op zoek.officielebekendmakingen.nl.
- Zorgverzekeringswet (2004) *Memorie van toelichting*. Kamerstuk 29763 nr. 3. Te vinden op zoek.officielebekendmakingen.nl.