

# Complex risicovereveningsmodel zorgverzekeraars blijft precies voorspellen

Om risicoselectie te voorkomen, worden zorgverzekeraars gecompenseerd voor verzekerden met een hogere kans op zorgkosten. Dat risicovereveningsmodel is steeds complexer geworden. Gaat dat ten koste van de statistische voorspelkracht?

## IN HET KORT

- Het risicovereveningsmodel wordt steeds complexer omdat er jaarlijks meer kenmerken aan worden toegevoegd.
- Geen van de bestaande kenmerken is overbodig om de voorspelkracht van het risicovereveningsmodel te optimaliseren.
- Overfitting is dus geen argument om de complexiteit van het risicovereveningsmodel te beperken.

**BENJAMIN VAN CASTEREN**  
Onderzoeker bij SEO

**MENNO POMP**  
Onderzoeker bij SEO

**JOERI ATHMER**  
Onderzoeker bij SEO

**WOUTER VERMEULEN**  
Hoofd cluster Zorg & Sociaal Domein bij SEO

Zorgverzekeraars zijn verplicht om iedereen te accepteren tegen dezelfde premie per zorgpolis. Omdat de verwachte zorgkosten per verzekerde sterk uiteenlopen, van enkele honderden euro's tot meer dan een half miljoen euro per jaar, ontvangen zorgverzekeraars voor elke verzekerde een op voorspelbare verschillen in zorgkosten gebaseerde vereveningsbijdrage. Dit risicovereveningssysteem is ontworpen om eerlijke concurrentie tussen zorgverzekeraars te bevorderen en risicoselectie te voorkomen.

Voor de risicoverevening maakt het systeem gebruik van een aantal kenmerken van de verzekerde, zoals leeftijd, geslacht en medische gegevens. Elk kenmerk is onderverdeeld in meerdere risicoklassen.

De risicoverevening is sinds de invoering in 1993 geleidelijk verbeterd, voornamelijk door nieuwe kenmerken aan het model toe te voegen en bestaande kenmerken verder uit te breiden. Figuur 1 laat zien hoe het aantal risicoklassen in de afgelopen jaren gestaag is toegenomen. In 2024 bestaat het model uit dertien kenmerken met in totaal 231 risicoklassen, terwijl er tien jaar geleden nog slechts acht kenmerken en 132 risicoklassen waren. Vlak voor de zomer heeft minister Dijkstra van Medische Zorg het risicovereveningsmodel voor 2025 nog verder uitgebreid met nieuwe risicoklassen voor het gebruik van langdurige zorg (Tweede Kamer, 2024).

Met de uitbreidingen wordt het model steeds complexer en daarmee minder goed uitvoerbaar. Het correct

toepassen van de regels en het beoordelen van de plausibiliteit van de uitkomsten worden steeds uitdagender voor het Zorginstituut Nederland (Hamstra et al., 2023).

Daarnaast kan de complexiteit leiden tot ongewenste en soms onvoorziene financiële prikkels (Kan et al., 2019; McGuire et al., 2023). Veel kenmerken zijn immers gebaseerd op het gebruik van specifieke zorg waardoor verzekeraars met hun inkoopbeleid kunnen sturen op het aanbod van die specifieke behandelingen.

De roep om het model te vereenvoudigen wordt dan ook steeds luider. In haar Kamerbrief over de vormgeving van het risicovereveningsmodel vraagt minister Dijkstra voor Medische Zorg nadrukkelijk om aandacht voor dit aspect (Tweede Kamer, 2024).

Naast deze uitvoeringsrisico's en ongewenste prikkelwerking is er mogelijk ook een statistisch argument tegen verdere uitbreiding van het model. Het is namelijk mogelijk dat het model, door extra kenmerken toe te voegen, zo sterk wordt afgestemd op historische gegevens dat dit ten koste gaat van de voorspelkracht op nieuwe, ongeziene data. Dit fenomeen staat bekend als *overfitting* (Webb, 2011). Omdat het risicovereveningsmodel voor 2025 gebaseerd is op zorgkostengegevens uit 2022, zou overfitting ertoe leiden dat het model minder goed corrigeert voor voorspelbare verschillen in zorgkosten in 2025. Hierdoor lopen verzekeraars meer financieel risico dan noodzakelijk.

Hoewel Amerikaans onderzoek laat zien dat overfitting in de risicoverevening geen denkbeeldig risico is (Kan et al., 2019), is dit voor Nederland nog niet onderzocht.

Figuur 1 laat wel zien dat de uitbreidingen in de afgelopen jaren weinig extra verklaringskracht hebben opgeleverd. Een substantiële toename van de R-kwadraat zoals tussen 2015 en 2017 is in de jaren daarna ondanks een gestage toename van risicoklassen niet meer gerealiseerd. Dit roept de vraag op in hoeverre deze extra risicoklassen daadwerkelijk hebben bijgedragen aan de voorspelkracht van het model. Als er overbodige kenmerken in het model zitten die niets toevoegen aan de voorspelkracht van het model, is er mogelijk ook in Nederland sprake van overfitting.

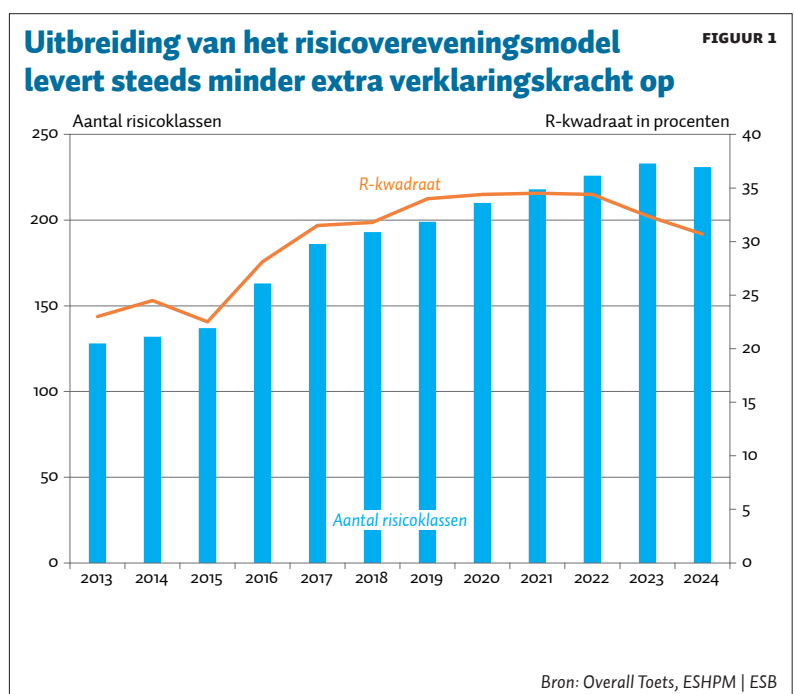
Aangezien de R-kwadraat enkel de verklaringskracht op historische data weergeeft en niets zegt over de voorspelkracht op ongeziene data, kunnen er uit deze figuur echter geen directe conclusies over overfitting worden getrokken. Bovendien wordt de ontwikkeling in de R-kwadraat

niet alleen bepaald door modelwijzigingen. De daling in de laatste twee jaren, waarin het model is geschat op zorguitgaven in 2020 en 2021, heeft bijvoorbeeld te maken met de coronapandemie.

In dit artikel onderzoeken we aan de hand van een methode uit de *machine learning* in hoeverre overfitting een probleem vormt voor het huidige risicovereveningsmodel en welke risicoklassen eventueel een minimale of negatieve bijdrage leveren aan de voorspelkracht van dit model.

### De LASSO

We onderzoeken de mogelijkheid van overfitting in het risicovereveningsmodel met de *Least Absolute Shrinkage and Selection Operator* (LASSO). De LASSO optimaliseert de set van risicoklassen met het oog op voorspelkracht door irrelevante of overbodige risicoklassen te verwijderen. De LASSO is daarmee nauw verwant aan kleinste-kwadratenmethode (OLS). OLS beschrijft de relatie tussen kenmerken en een uitkomstvariabele zo goed mogelijk, maar dat kan leiden tot slechtere voorspellingen als het aantal kenmerken te groot wordt. LASSO biedt hier een oplossing door een straf toe te voegen voor de grootte van de coëfficiënten in het model. Deze straf zorgt ervoor dat sommige coëfficiënten letterlijk nul worden, waardoor die kenmerken niet meer meedoen in het model. Dit proces wordt kenmerkselectie genoemd. Op deze manier houdt LASSO alleen de meest relevante kenmerken over en verwijdert het de ruis van minder belangrijke of irrelevante kenmerken.



Wij passen de LASSO toe op het databestand dat gebruikt wordt voor de schatting van het somatische risicovereveningsmodel voor 2024 (Cattell et al., 2023). Dit bestand bevat ongeveer achttien miljoen verzekerden, hun zorgkosten in 2021 en de kenmerken die deze zorgkosten in het model verklaren. We nemen de volledige verzameling kenmerken uit de somatische risicoverevening mee.

## De LASSO biedt geen verbetering ten opzichte van OLS

TABEL 1

Methode	Aantal behouden klassen	R-kwadraat bij het trainen	R-kwadraat bij het testen
OLS	207	30,6%	30,8%
LASSO	203	29,7%	29%

Noot:  $N = 8.787.213$  (trainingsdata),  $N = 8.787.214$  (testdata)

Dit leidt tot 207 risicoklassen. Net zoals in het risicovereveningsmodel vormen de zorgkosten de afhankelijke variabele in onze analyse.

Om het model op overfitting te evalueren, gebruiken we een opdeling van het databestand in twee even grote deelbestanden. Eén deelbestand wordt gebruikt om het model te schatten (het trainingsbestand), terwijl het andere bestand uitsluitend dient voor validatie (het testbestand). Door het model te testen op een dataset die niet is gebruikt bij de schatting, kunnen we de voorspelkracht op ongeziene data beoordelen. Een hogere R-kwadraat op het testbestand wijst op betere voorspellingsprestaties.

Om te bepalen hoe hoog de straf voor de grootte van de coëfficiënten in het LASSO-model precies moet zijn, maken we gebruik van kruisvalidatie. Hierin volgen we de aanpak van Rose (2016). Dit betekent dat we het trainingsbestand opnieuw opsplitsen in vijf deelbestanden en de hoogte van de straf en de coëfficiënten van het model op verschillende combinaties van deze deelbestanden schatten. Uiteindelijk kiezen we de hoogte van de straf die op alle vijf de deelbestanden tot de beste modelprestaties leidt. Zo verkleinen we het risico op overfitting dat zou kunnen ontstaan door eigenaardigheden als gevolg van een willekeurige uitsplitsing van het trainingsbestand.

Ter vergelijking schatten we ook een OLS-model op het trainingsbestand, dat we op dezelfde wijze valideren met het testbestand. Overfitting wordt zichtbaar wanneer het LASSO-model een groot aantal overbodige risicoklassen weglaat en wanneer de R-kwadraat van het LASSO-model op het testbestand veel hoger is dan die van het OLS-model. Deze combinatie wijst erop dat het model historische data te specifiek heeft verklaard, zonder dat dit leidt tot betere voorspellingen op nieuwe data.

Met deze aanpak kunnen we niet alleen vaststellen of overfitting een probleem is, maar ook inzicht krijgen in welke kenmerken een minimale of negatieve bijdrage leveren aan de voorspelkracht van het risicovereveningsmodel.

## Resultaten

De resultaten in tabel 1 wijzen niet op overfitting. LASSO blijkt slechts vier van de 207 risicoklassen te verwijderen. Het gaat om de klassen *Bijstandsgerechtigden 35–44 jaar*, *Bijstandsgerechtigden 45–54 jaar* en *Zelfstandigen 65–69 jaar* binnen het kenmerk Aard van Inkomen (AVI) en de klasse *Blijvend in Wlz-instelling 70–79 jaar* binnen het kenmerk Personen per Adres (PPA). De andere 203 klassen zijn nodig om de voorspelkracht van het model te optimaliseren. Bovendien vallen de vier weggevallen klassen binnen

kenmerken die uit veel meer risicoklassen bestaan. Het is dus niet zo dat er hele kenmerken weggelaten kunnen worden, zonder dat dit ten koste gaat van de voorspelkracht van het model.

Afgemeten aan de R-kwadraat presteert het OLS-model bij zowel het schatten als het valideren beter dan de LASSO. Door risicoklassen weg te laten, verslaat de LASSO het OLS-model dus niet op voorspelkracht. Dat de R-kwadraat van het OLS-model op het testbestand zelfs iets hoger uitkomt dan bij het schatten wijst er ook op dat overfitting geen rol speelt: dit model presteert niet slechter op ongeziene data dan op de data waarop het is geschat.

## Conclusies

Ondanks de groei in het aantal kenmerken en risicoklassen in het risicovereveningsmodel en de afnemende meeropbrengst in verklarende kracht, is er op dit moment geen sprake van overfitting. Uit onze toepassing van de LASSO-methode blijkt dat het model geen overbodige kenmerken bevat. De LASSO-analyse verwijdert slechts een beperkt aantal risicoklassen om de voorspelkracht te optimaliseren en dit biedt geen grondslag voor het verwijderen van hele kenmerken, die uit veel meer risicoklassen bestaan. Bovendien blijkt het OLS-model, in vergelijking met het LASSO-model, beter te presteren op ongeziene data.

Hoewel overfitting dus in potentie een statistisch argument vormt om de complexiteit van het risicovereveningsmodel te beperken, blijkt dit in de praktijk voor Nederland nog niet relevant. Dit laat onverlet dat de andere argumenten op het terrein van uitvoerbaarheid en ongewenste prikkelwerking van kracht blijven. Het kritisch evalueren van de modelcomplexiteit blijft daarom van onverminderd belang.

## Literatuur

- Cattel, D., F. Eijkenaar, M. Oskam et al. (2023) *Onderzoek risicoverevening 2024: Overall Toets*. Erasmus School of Health Policy & Management, Rapport, 3 oktober. Te vinden op [www.rijksoverheid.nl](http://www.rijksoverheid.nl).
- Hamstra, G., S. Borg, P. Suurenbroek en P. Stam (2023) *Complexiteit in de uitvoering van de risicoverevening*. Equalis Rapport, 26 juni. Te vinden op [www.zorginstituutnederland.nl](http://www.zorginstituutnederland.nl).
- Kan, H.J., H. Kharrazi, H.-Y. Chang et al. (2019) Exploring the use of machine learning for risk adjustment: A comparison of standard and penalized linear regression models in predicting health care costs in older adults. *PLoS ONE*, 14(3), e0213258.
- McGuire, T.G., A.L. Zink en S. Rose (2021) Improving the performance of risk adjustment systems: Constrained regressions, reinsurance, and variable selection. *American Journal of Health Economics*, 7(4), 497–521.
- Rose, S. (2016) A machine learning framework for plan payment risk adjustment. *Health Services Research*, 51(6), 2358–2374.
- Tweede Kamer (2024) *Kamerbrief ex-ante vormgeving risicovereveningsmodel 2025*, 3847848-1067510-Z, 17 juni.
- Webb, G.I. (2011) Overfitting. In: C. Sammut en G.I. Webb (red.), *Encyclopedia of Machine Learning*. Boston, MA: Springer, p. 744.