

Arbeidsmarkt- en gezondheidsrisico's stapelen zich op bij dezelfde personen

Tegenslagen zoals ontslag of ziekte houden vaak met elkaar verband, maar hoe groot is deze samenhang? Met behulp van 'machine learning' kunnen we de totale kansverdeling op tegenslagen op het gebied van arbeid en gezondheid in beeld brengen. Dit biedt nieuwe mogelijkheden voor preventief beleid.

IN HET KORT

- Arbeidsmarkt- en gezondheidsrisico's zijn ongelijk verdeeld: slechts een kleine groep heeft een aanzienlijke kans.
- Risico's stapelen: de top één-procent met het hoogste zorgkostenrisico heeft vier keer meer kans op uitkeringsafhankelijkheid.
- Eerdere tegenslagen zijn indicatief voor toekomstige tegenslagen, wat een aanknopingspunt biedt voor gericht beleid.

EMILE CAMMERAAT

Programmaleider Risico & Regulering bij het Centraal Planbureau (CPB)

BRINN HEKKELMAN

Wetenschappelijk medewerker bij het CPB

PIM KASTELEIN

Wetenschappelijk medewerker bij het CPB

SUZANNE VISSERS

Wetenschappelijk medewerker bij het CPB

Het ministerie van SZW heeft het onderliggende CPB-onderzoek (Cammeraat et al., 2023) medegefinancierd.

Tegenslagen in mensenlevens staan doorgaans niet op zichzelf. Ziekte kan leiden tot baanverlies, en schulden kunnen leiden tot mentale gezondheidsproblemen (García-Gómez et al., 2013; Roos et al., 2021).

Er bestaan verschillende studies die kwetsbare groepen in Nederland in kaart brengen (De Klerk et al. 2023; Van Hoenselaar et al., 2023), en initiatieven zoals de Kansentlas (SEO Economisch Onderzoek) en de KansentKaart (Erasmus Universiteit Rotterdam). Wat echter nog niet goed is onderzocht, is de vraag in hoeverre het mogelijk is om dergelijke negatieve gebeurtenissen van tevoren te zien aankomen. Is een tegenslag volledig willekeurig, of is er enige mate van voorspelbaarheid waarop mogelijk kan worden geanticipeerd? En hoe verhouden risico's op verschillende tegenslagen zich tot elkaar?

In dit artikel schatten we de voorspelbare factor die voorafgaat aan de tegenslagen. Deze factor kan geïnterpreteerd worden als het risicotype van een individu (Mueller en Spinnewijn, 2023). Het gaat hierbij om een 'policy prediction problem', wat inhoudt dat we geïnteresseerd zijn in hoe mensen er in de toekomst voor staan, gegeven hun huidige situatie (Kleinberg et al., 2015; Mullainathan en Spiess, 2017). Hiervoor is het nodig dat we accuraat inschattingen kunnen maken, maar causaliteit is daarbij niet vereist.

Van realisaties achteraf naar risico's vooraf

Op basis van administratieve gegevens van het Centraal Bureau voor de Statistiek hebben we een dataset samengesteld met ruim 500 variabelen voor de periode 2013–2018, met daarin informatie over demografische en sociaal-economische kenmerken, en arbeid en gezondheid voor miljoenen Nederlanders.

Vervolgens hanteren we twee schokdefinities: ten eerste, de primaire inkomensbron van een persoon wordt een uitkering (bij werkloosheid, arbeidsongeschiktheid, ziekte of bijstand), en ten tweede, de gezondheidskosten van een persoon stijgen met meer dan 5.000 euro in een kalenderjaar. In de onderliggende onderzoekspaper (Cammeraat et al., 2023) verkennen we nog tien alternatieve schokdefinities binnen de domeinen van arbeid en zorg.

Wij maken gebruik van *machine learning* omdat we primair geïnteresseerd zijn in het schatten van de kansen op nadelige gebeurtenissen, en niet zozeer in de onderliggende redenen en verklarende variabelen. *Machine learning* is geschikt voor het maken van dergelijke inschattingen, omdat het in staat is om grote hoeveelheden data te verwerken en complexe interacties mee te nemen. Met behulp van deze methodiek bepalen we, op basis van recente persoonsgegevens, voor elk individu in ieder jaar de kans op een schok in het komende jaar. Voor elke schokdefinitie wordt een model gemaakt dat is getraind en getest op twee verschillende datasets, elk bestaande uit zo'n twee miljoen verschillende 'individu-jaar-combinaties'.

Het daadwerkelijke risico dat mensen lopen op een tegenslag is een niet observeerbare (latente) variabele. Onze methodiek levert hier per individu-jaar-combinatie een inschatting van. Vervolgens controleren we of onze schattingen gemiddeld genomen overeenkomen met de daadwerkelijke schokrealisaties in de data, en bestuderen we op groepsniveau de samenhang tussen kansschattingen voor de verschillende tegenslagen.

Voorspelbaarheid en stapeling van risico's

Arbeidsmarktrisico's zijn accuraat te voorspellen (figuur 1). In het hoogste percentiel schat het model de gemiddelde kans op de uitkeringschok op zo'n zeventig procent. De daadwerkelijke prevalentie van die schok bij de individuen in dat percentiel is ongeveer 67 procent, ten opzichte van een populatiegemiddelde van zo'n 2,3 procent.

Uit de figuur valt een aantal zaken af te leiden. Ten eerste concentreert het risico zich bij een kleine groep mensen. Ongeveer tachtig procent van de mensen heeft een risico lager dan gemiddeld; en ongeveer vijftig procent loopt vrijwel geen risico. Het risico voor de individuen in de top drie-procent van de risicoverdeling is tien keer zo groot als het populatiegemiddelde, voor de top één-procent is dit zelfs een factor dertig. Daarnaast valt op dat de voorspellingen en realisaties niet ver uit elkaar liggen. Dit toont aan dat ons model daadwerkelijk het risico op een schok nauwkeurig kan inschatten. Overigens zijn de rijke data hierbij van essentieel belang: een robuustheidstest waarbij we het model hebben getraind met slechts een aantal basale persoonskenmerken liet vrijwel geen voorspelkracht zien.

Vervolgens gaan we een stap verder en demonstreren we dat een verhoogd risico op de ene schok gepaard gaat met een verhoogd risico op de andere schok. Figuur 2 geeft een weergave van het positieve verband tussen de twee schokken.

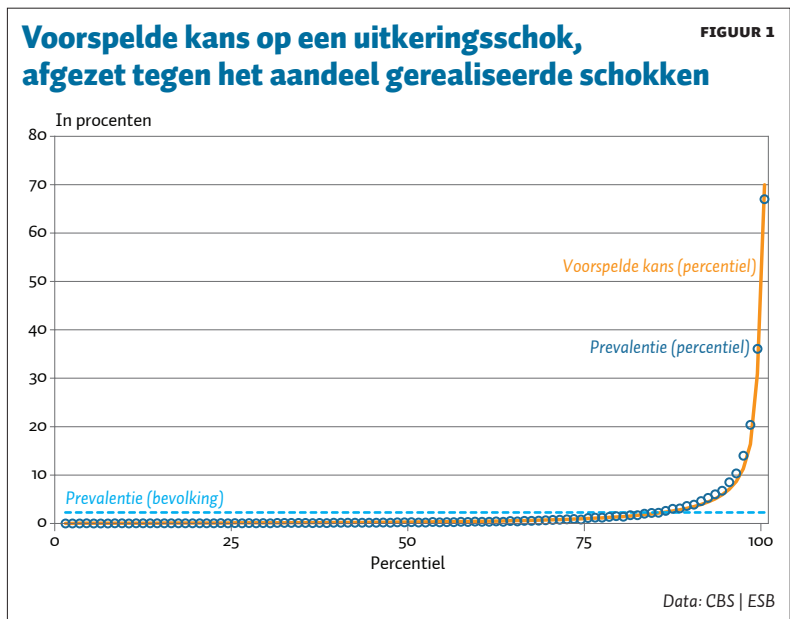
De voorspellingen en realisaties liggen weer dicht bij elkaar. Daarnaast valt het positieve verband tussen de twee schokken direct op: een hoger risico op een gezondheidsschok gaat gepaard met een hoger risico op een arbeidsschok. De groep personen met de één procent hoogste kans op de zorgkostenschok, heeft ook een vier keer zo hoge kans om afhankelijk te worden van een uitkering ten opzichte van het bevolkingsgemiddelde. Er is overigens over de gehele risicoverdeling sprake van een positief verband tussen de zorgkostenschok en uitkeringsschok. Een stapje naar een hoger percentiel in de risicoverdeling van de gezondheidsschok impliceert daarmee ook een stijging in de risicoverdeling van de arbeidsschok, ongeacht de oorspronkelijke positie in de risicoverdeling van de gezondheidsschok. Juist door de gehele risicoverdeling in kaart te brengen, in plaats van slechts te focussen op de realisaties, hebben we deze percentielgroepen kunnen construeren en dit fenomeen inzichtelijk kunnen maken.

Hetzelfde patroon komt naar voren voor alle andere schokcombinaties die in het onderzoekspaper zijn bestudeerd, zowel binnen als tussen de domeinen arbeid en zorg. Dit wijst erop dat de samenloop van risico's een wijdverspreid fenomeen is.

Terwijl in figuur 2 inschattingen gemaakt worden voor schokken in hetzelfde jaar, kan er ook gekeken worden naar opeenvolgende schokken. Gegeven dat een individu in het voorafgaande jaar een grote stijging van zorgkosten heeft gehad, is de kans om afhankelijk te worden van een uitkering bijna drie keer zo groot. Positieve verbanden van deze orde grootte komen in de analyse naar voren voor een breed scala aan schokdefinities. Zo is de kans dat iemand begint met het innemen van medicatie tegen psychische klachten twee keer zo groot als die in het jaar daarvoor gestopt is met het betalen van de zorgverzekeringspremies – een indicator voor het hebben van financiële problemen.

Mogelijkheden voor preventief beleid

De sterke ongelijke verdeling van risico's betekent dat de meeste mensen amper risico lopen op ziekte of werkloosheid, terwijl een paar procent van de bevolking een hoge voorafkans heeft op beide schokken. In plaats van dat



iedereen af en toe een tegenslag heeft, zijn deze tegenslagen dus geconcentreerd bij een klein deel van de bevolking.

Door de sterke mate waarin de risico's samenlopen, is er ook sprake van een effect in andere domeinen. Zo kan het verminderen van het risico op werkloosheid ook het risico op latere uitgaven voor geestelijke gezondheidszorg verminderen.

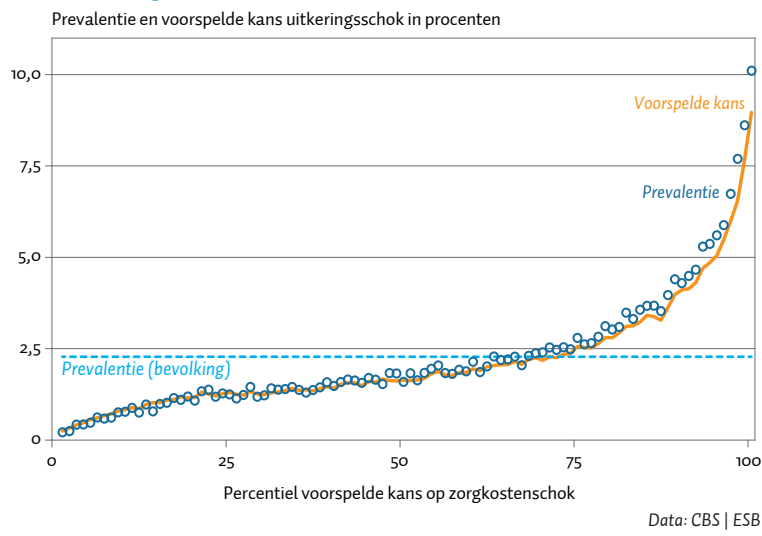
De sterke ongelijkheid in de blootstelling aan schokken kan een reden zijn om de kleine groep mensen met de hoogste kans op tegenslagen extra te steunen om zo de bestaanszekerheid te vergroten, bijvoorbeeld met aanvullende inkomenssteun.

De vraag is of het huidige sociale zekerheidsstelsel voldoende rekening houdt met sterk ongelijke kansen en stapelende risico's. Het huidige stelsel beschouwt de schokken los van elkaar: we hebben weliswaar toeslagen en uitkeringen voor losse zaken, maar nemen onvoldoende de samenhang in ogenschouw. Daarnaast is het stelsel reactief en niet proactief: er wordt gereageerd nadat een schok heeft plaatsgevonden.

Het is in principe mogelijk om groepen individuen aan te wijzen met een grotere kans op een arbeidsmarkt- en/of gezondheidsschok. Preventief beleid gericht op deze groepen zou effectiever kunnen zijn dan het aanpakken van de gevolgen van tegenslagen. Een voorbeeld is om werkloosheid te voorkomen door de arbeidsmarktpositie van personen met een hoog risico op baanverlies te verbeteren. Dit geniet mogelijk de voorkeur boven het verstrekken van werkloosheidsuitkeringen na daadwerkelijk baanverlies.

De vraag is hoe we effectief beleid kunnen toespitsen op de mensen die het meeste risico lopen. Men zou in de verleiding gebracht kunnen worden om de individuele risicoschattingen van onze modellen direct te gebruiken, maar dit is niet mogelijk. Modellen zoals wij gebruikt hebben voor ons onderzoek, zijn in de praktijk moeilijk in te zetten, aangezien de benodigde data niet tijdig verzameld kunnen worden om risico-inschattingen te maken van schokken voordat die zich voordoen. Naast deze praktische overweging bestaan er door onder meer de toeslagen-

De kans op een uitkeringsschok gegeven de kans op een zorgkostenschok FIGUUR 2



Kans op uitkeringsschok voor verschillende subpopulaties TABEL 1

	Percentage individuen met uitkeringsschok	Fractie van populatie, in procenten
Totale bevolking	2,3	100,0
Top één-procent-risico	67,0	1,0
Huurhuis met huurtoeslag (1)	7,6	5,6
Laag opleidingsniveau (2)	5,5	9,5
Tijdelijk contract (3)	4,9	20,0
Achtergrondkenmerken (1)+(2)+(3)	14,3	0,5
Problematische schuld (t-1)	10,2	0,4
Stijging zorgkosten (t-1)	6,4	2,3
Mentale zorg medicatie (t-1)	7,6	1,7

Data: CBS | ESB

affaire twijfels over de wenselijkheid en transparantie van een dergelijke datagestuurde aanpak.

Daarentegen is het wel mogelijk om preventief beleid te richten op subgroepen. Dit vereist inzicht in het risico op schokken voor verschillende subgroepen.

Tabel 1 geeft voor verschillende subgroepen aan hoe vaak de schok om afhankelijk van een uitkering te worden voorkomt (prevalentie). In het ene uiterste geval, waarbij beleid volledige ongericht zou zijn, vindt bij 2,3 procent van de individuen daadwerkelijk een schok plaats. In het andere uiterste geval, waarbij de subgroep bestaat uit alleen de 1 procent individuen die in ons model het hoogste risico hebben, ondergaat 67 procent van de individuen in deze groep inderdaad een schok.

Van de mensen die in een problematische schuldenpositie zijn geraakt in het voorgaande jaar (hier gedefinieerd als het niet meer betalen van de zorgverzekeringspremie), belandt 10,2 procent het jaar daarop in een uitkering. In de groep die in het voorgaande jaar is gestart met het innemen van medicatie tegen psychische klachten, is dit 7,6 procent. Tegenslagen in het verleden zijn dus informatief voor toe-

komstige tegenslagen in een ander domein, en ze presteren minstens net zo goed als vaak gebruikte (combinaties van) persoonskenmerken. Schokrealisaties uit het verleden zouden daarom kunnen worden meegenomen in de selectie van doelgroepen voor gericht preventief beleid.

Conclusie

Het is mogelijk om de risico's op arbeidsmarkt- en gezondheidsschokken, en de samenhang daartussen, accuraat in beeld te brengen. De aanwezigheid van een aanzienlijke samenloop tussen risico's impliceert dat gericht preventiebeleid, dat proactief meerdere kwetsbaarheden aanpakt, veelbelovend is.

Gegeven dat het praktisch onhaalbaar en mogelijk onwenselijk is om beleid te richten op basis van voorspellingen op individueel niveau, kan het verstandig zijn om preventief beleid meer te richten op subgroepen met een hoge kans op een schok. Tegenslagen uit het verleden en verschillende domeinen zijn hierbij een waardevol aanvullend criterium voor beter gericht beleid.

Idealiter wordt preventief beleid niet alleen gericht op mensen die een grote kans op een tegenslag hebben, maar ook op mensen die moeizaam herstellen van een tegenslag. In een vervolgstudie naar weerbaarheid brengen we het ongelijke herstel van tegenslagen in beeld en relateren we dit aan de vooraf geschatte risico's.

Literatuur

- Cammeraat, E., B. Hekkelman, P. Kastelein en S. Vissers (2023) *Predictability and (co-)incidence of labor and health shocks*. CPB Discussion Paper, 453.
- García-Gómez, P., H. van Kippersluis, O. O'Donnell en E. van Doorslaer (2013) Long-term and spillover effects of health shocks on employment and income. *The Journal of Human Resources*, 48(4), 873–909.
- Hoenselaar, F. van, G. Eijnsink en N. Rupert (2023) *Kwetsbaarheid en veerkracht van Nederlandse huishoudens*. DNB Occasional Studies, 21-1.
- Kleinberg, J., J. Ludwig, S. Mullainathan en Z. Obermeyer (2015) Prediction policy problems. *The American Economic Review*, 105(5), 491–495.
- Klerk, M. de, E. Eggink, I. Plaisier en K. Sadiraj (2023) *Zicht op zorgen*. Sociaal en Cultureel Planbureau.
- Mueller, A.I. en J. Spinnewijn (2023) *The nature of long-term unemployment: Predictability, heterogeneity and selection*. NBER Working Paper, 30979.
- Mullainathan, S. en J. Spiess (2017) Machine learning: An applied econometric approach. *The Journal of Economic Perspectives*, 31(2), 87–106.
- Roos, A.-F., M. Diepstraten en R. Douven (2021) *When financials get tough, life gets rough? Problematic debts and ill health*. CPB Discussion Paper, 428.