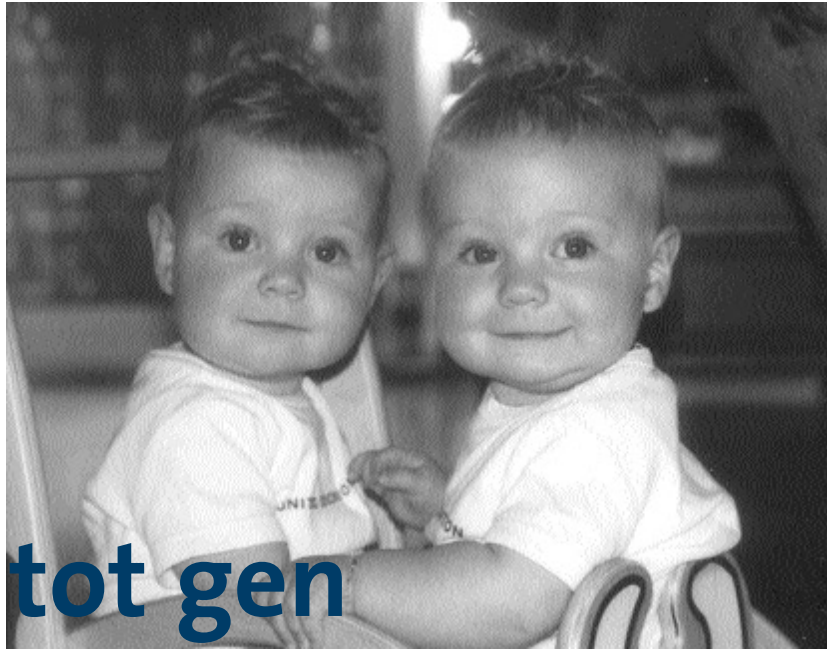


Van SEM tot gen



Toepassing van structurele vergelijkingsmodellen in de gedragsgenetica

Eeneiige tweelingen zijn individuen met een identieke genetische aanleg, terwijl twee-eiige tweelingen slechts een deel van hun genen gemeen hebben. Dit gegeven heeft geleid tot de ontwikkeling van een reeks statistische modellen voor het bepalen van de erfelijkheid van zowel fysieke als psychologische kenmerken. De takken van wetenschap die zich hiermee bezighouden zijn de gedragsgenetica en de genetische epidemiologie. Hun primaire doel is te onderzoeken of, en zo ja in welke mate, de oorzaken van verschillen in gedrag, of in ziekte en gezondheid, tussen mensen kunnen worden toegeschreven aan erfelijke factoren en omgevingsinvloeden. *Structural equation modeling* (maar zeker ook de komst van het computerprogramma Mx van Mike Neale) heeft het specificeren en testen van zowel simpele als meer complexe genetische modellen aanzienlijk vereenvoudigd. De combinatie van *structural equation modeling* en grote datasets, met informatie over tweelingen en ouders zoals die bijvoorbeeld aanwezig is op de Afdeling Biologische Psychologie van de Vrije Universiteit, is dan ook veel belovend.

Persoonskenmerken clusteren binnen families. Dit gegeven is op zich zelf nog niet voldoende om de conclusie te trekken dat er sprake is van erfelijke of genetische invloeden. Het is immers mogelijk dat de door de familieleden gedeelde omgeving heeft bijgedragen aan de gelijkenis van verschillende gezins- en familieleden. Zonder extra informatie zijn genetische factoren niet te scheiden van de invloeden uit de gedeelde omgeving.

Het adoptiedesign

Een veel gebruikt design om genetische factoren te scheiden van omgevingsinvloeden is het adoptiedesign, waarin geadopteerde kinderen worden vergeleken met zowel hun biologische als niet-biologische ouders. Geadopteerde kinderen delen alleen hun genen en prenatale omgeving met hun biologische ouders, terwijl ze de omgeving alleen delen met hun adoptie-ouders. Als adoptiekinderen op een bepaald kenmerk meer lijken op hun biologische ouders dan op hun niet-biologische ouders dan is dit een sterke aanwijzing voor de invloed van erfelijke factoren op dit kenmerk. Andersom, als ze meer gelijkenis vertonen met hun niet-biologische ouders dan krijgen omgevingsinvloeden meer gewicht. Met dit design zijn enkele belangrijke successen geboekt. Zo heeft het onderzoek geleid tot een keerpunt in het onderzoek naar het ontstaan van schizofrenie, een ziekte waarvan men aannam dat ze volledig veroorzaakt werd door omgevingsfactoren. De resultaten van Heston (1966) lieten zien dat 11% (5 van 47) van de geadopteerde kinderen waarvan de biologische moeder als schizofreen was gediagnosticeerd, zelf ook schizofrenie ontwikkelde. Een percentage dat aanzienlijk hoger is dan dat van 0% (0 van de 50) in de controlegroep van adoptiekinderen waarvan de biologische ouders geen schizofrenie hadden. Dit resultaat betekent dat de familiale over-

eenkomst met betrekking tot schizofrenie een genetische oorsprong heeft. Adoptiestudies in Scandinavië lieten bovendien zien dat opgroeien in een familie waarin schizofrenie voorkomt niet leidt tot een verhoogd risico.

De allereerste adoptiestudie stamt uit 1924 (Theis, 1924) en beschrijft een onderzoek naar de erfelijkheid van intelligentie, in een tijd dat adoptie nog relatief veel voorkwam. In de huidige maatschappij komt adoptie aanzienlijk minder vaak voor dan vroeger. Verder neemt de toepasbaarheid van het adoptiedesign af doordat er in de regel sprake is van selectieve plaatsing van kinderen in adoptiegezinnen. Adoptiegezinnen worden vaak gematched op relevante kenmerken van de biologische ouders, wat de scheiding van erfelijke factoren en omgevingsinvloeden bemoeilijkt. Deze en andere factoren hebben er toe geleid dat het adoptiedesign maar beperkt toepasbaar is en de nadruk binnen de gedragsgenetica is komen te liggen op methoden met bredere toepasbaarheid. Een van deze methoden betreft het tweelingdesign.

Het tweelingmodel

Eeneiige tweelingen, of monozygoten (MZ)¹, vormen een genetisch identiek paar. Dit betekent dat zij al hun genen gemeenschappelijk hebben. Ook hebben zij vanzelfsprekend een deel van hun omgeving gemeenschappelijk indien ze samen zijn opgevoed. De mate waarin MZ tweelingen van elkaar verschillen kan dus alleen worden toegeschreven aan unieke, niet gedeelde omgevingsinvloeden (inclusief een eventuele meetfout). Twee-eiige tweelingen, of dizygoten (DZ), delen hun omgeving ook, maar hebben net als andere eerstegraads familieleden gemiddeld 50% van hun genen gemeenschappelijk². De mate waarin DZ van elkaar verschillen kan dus zowel worden

toegeschreven aan genetische factoren als aan unieke, niet gedeelde omgevingsinvloeden. De redenatie is nu dat als genetische factoren van invloed zijn op een bepaald persoonskenmerk, MZ tweelingen een grotere gelijkenis vertonen op dit persoonskenmerk dan DZ tweelingen.

Afhankelijk van welke familieleden gegevens bekend zijn kan er een decompositie plaatsvinden van de totale fenotypische variantie (variantie van de geobserveerde kenmerken) in variantie ten gevolge van additieve genetische effecten (a^2), interactieve genetische effecten of dominantie (d^2), gedeelde omgevingseffecten (c^2), en unieke omgevingseffecten (e^2). Simultane schatting van de vier variantiecomponenten is uitgaande van het standaard tweeling design echter niet mogelijk, aangezien dit model niet geïdentificeerd is. Wel kan de passing van een ADE-model worden vergeleken met de passing van een ACE-model. Uit zowel empirisch als theoretisch onderzoek blijkt dat voor veel concepten in de psychologie de omgevingseffecten van groter belang zijn dan dominantie; dit impliceert een ACE-model. In de rest van dit verhaal zullen we ons daarom richten op de schatting van de A, C en E componenten. Een centrale statistische grootte in de gedragsgenetica is de erfelijkheid ('heritability' of h^2), gedefinieerd als de proportie fenotypische variantie die verklaard wordt door genetische verschillen tussen mensen. Bij afwezigheid van dominantie is de erfelijkheid gelijk aan $a^2/(a^2 + c^2 + e^2)$.

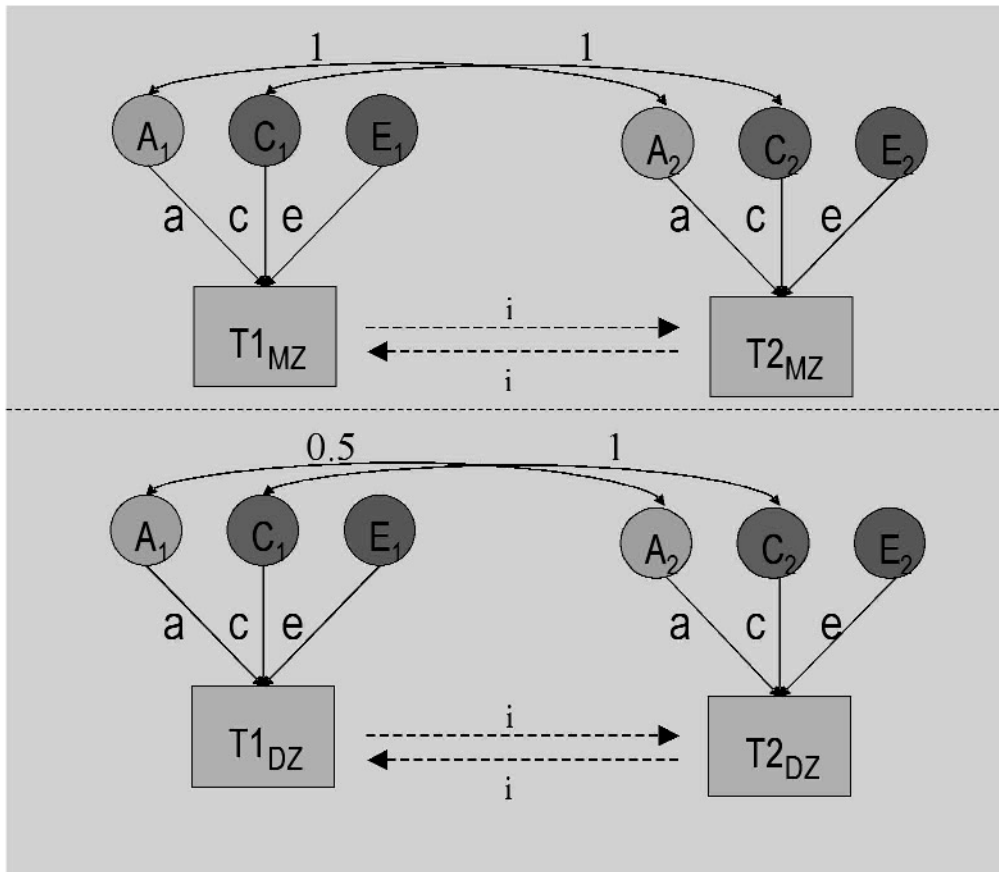
Een eerste schatting van de verschillende variantiecomponenten, uitgaande van een ACE model, kan worden verkregen met behulp van de fenotypische correlatie binnen MZ en DZ paren zoals staat vermeld in Tabel 1. De puntschattingen zoals die worden verkregen met de formules in Tabel 1 betreffen echter slechts een berekening van de grootte van de parameters. Met behulp van Structural Equation Modeling (SEM), uitgevoerd bijvoorbeeld met de computerprogramma's LISREL (Jöreskog en Sörbom) of Mx (Neale et al), kunnen additioneel ook schattingen van de standaardfouten en betrouwbaarheidsintervallen worden verkregen, alsmede van de overall fit van het model door middel van een likelihood-ratio test. Uiteraard onder de voorwaarde dat de aannames die aan SEM ten grondslag liggen niet zijn geschonden. SEM is een verdere uitwerking van de door Wright (1921) ontwikkelde padanalyse.

De formules in Tabel 1 houden er onder meer geen rekening mee dat de erfelijkheid bij mannen en vrouwen vaak niet gelijk is. SEM is bij uitstek geschikt voor analyseren van seksespecifieke modellen en allerlei multivariate uitbreidingen van het model in Tabel 1.

Een voorbeeld van een paddiagram van een 'ACE' model voor het klassieke tweelingdesign is te vinden in Figuur 1. We zullen deze figuur hier kort toelichten.

$r_{mz} = a^2 + c^2$	$r_{dz} = \frac{1}{2} a^2 + c^2$
$a^2 = 2(r_{mz} - r_{dz})$	
$c^2 = 2r_{dz} - r_{mz}$	$e^2 = 1 - r_{mz}$

Tabel 1: Berekening van gestandaardiseerde variantie componenten op basis van fenotypische correlaties r_{mz} en r_{dz} in het ACE model

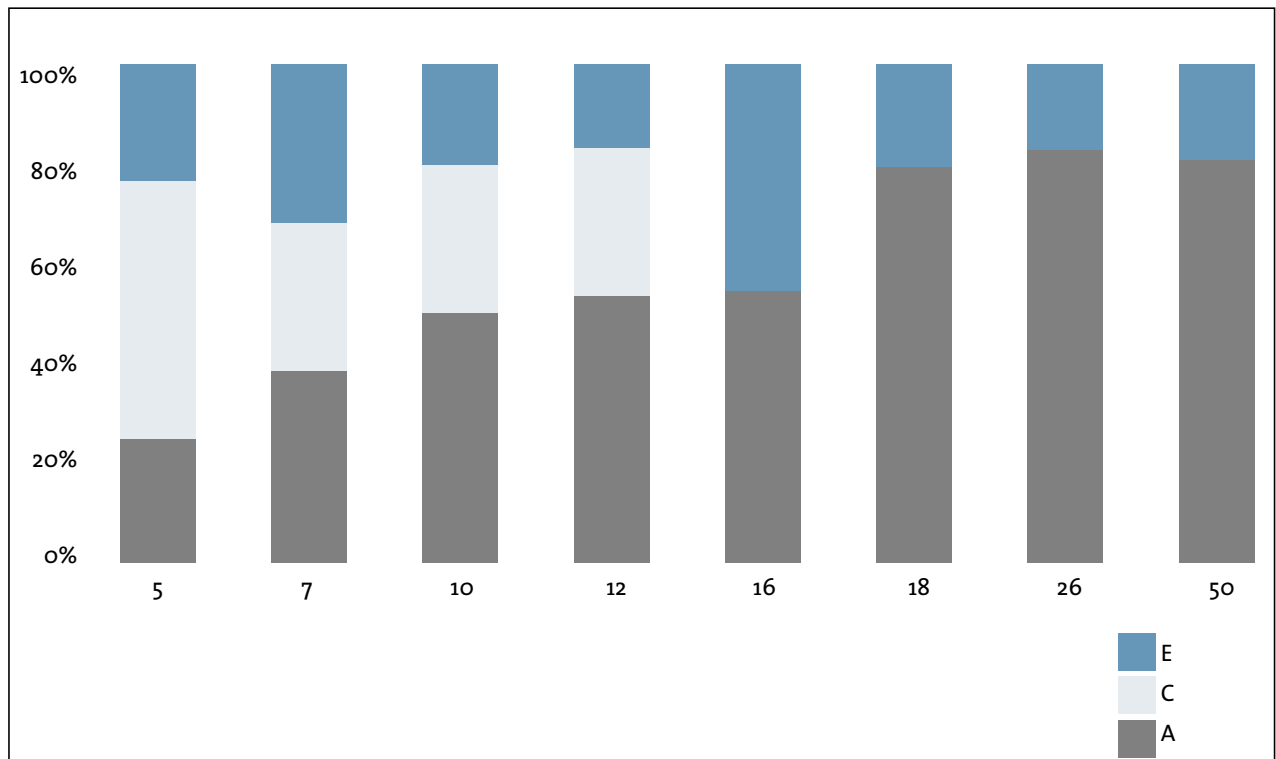


Figuur 1: Pad-diagram van het ACE model.

Figuur 1 betreft de decompositie van de fenotypische variantie van tweeling 1 en 2 (T_1 en T_2) in de drie componenten $A_{1(2)}$, $C_{1(2)}$ en $E_{1(2)}$, gerepresenteerd door gestandaardiseerde latente variabelen in het padmodel. De factorladingen (a , c en e) worden geschat en het kwadraat van de factorladingen geeft de variantie die verklaard door het overeenkomstige effect. Het model wordt geschat als tweegroepenmodel, waarbij een groep de MZ betreft, en de andere de DZ. C is gedefinieerd door middel van de restrictie van de correlatie tussen T_1 en T_2 op de waarde 1 in beide groepen, A door restrictie van de correlatie bij MZ op 1 en bij DZ op 0.5 (zie boven), en E blijft ongecorrleerd in beide groepen en bevat dus naast de unieke omgeving ook een eventuele meetfout. Verder worden de factorladingen (a , c en e) zowel binnen als tussen

de groepen aan elkaar gelijk gesteld. Indien er ook een groep met geadopteerde tweelingen zou zijn opgenomen in het model, dan zou daar de waarde van de C geresliceerd worden op de waarde nul.

Normaliter is een tweegroepenmodel met drie latente variabelen per geobserveerde variabele niet geïdentificeerd. Identificatie wordt hier echter verkregen door middel van de restricties van parameters zowel binnen als tussen de groepen. Uit het paddiagram kan worden afgeleid dat de modelgeïmpliceerde variantie van zowel T_1 als T_2 in beide groepen gelijk is aan: $a^2+c^2+e^2$. De modelgeïmpliceerde covariantie tussen T_1 en T_2 bij de MZ is gelijk aan: a^2+c^2 ; terwijl deze bij DZ gelijk is aan: $\frac{1}{2}a^2+c^2$. Met deze parameters kunnen de modelgeïmpliceerde covariantie-matrices worden opgesteld, die vervolgens worden vergeleken



Figuur 2: Erfelijkheid van intelligentie op verschillende leeftijden. Leeftijd 5 tot 12 betreffen herhaalde metingen binnen dezelfde kinderen. Leeftijd 16-50 betreffen onafhankelijke steekproeven.

met de geobserveerde covariantie-matrices³. Het model kan worden uitgebreid met een wederkerige relatie tussen de tweelingen, gerepresenteerd door de parameter i . Deze parameter representeert de wederzijdse invloed van tweelingen op elkaar op fenotypisch niveau en heeft tot gevolg dat de totale varianties voor MZ en DZ van elkaar gaan verschillen. Deze parameter blijkt met name van belang bij variabelen als rookgedrag, en taalgebruik, maar minder bij bijvoorbeeld intelligentie en lengte. Tevens kan i een indicatie zijn voor de aanwezigheid van zogenaamde 'rater-bias'.

In een recente analyse van de intelligentiescores op de RAKIT van 12-jarige scholieren met behulp van het ACE model werd een erfelijkheid gevonden van 51% (Bartels, Rietveld, Van Baal & Boomsma, 2002). De gemeenschappelijke omgeving verklaarde 33% van de verschillen in intelligentie

tussen scholieren en de overige 15% konden worden toegeschreven aan de unieke omgeving of meetfouten. Deze resultaten zijn consistent met bestaande kennis op het gebied van de erfelijkheid van cognitieve vaardigheden. Tevens blijkt uit onderzoek dat de erfelijkheid van intelligentie toeneemt met leeftijd. Dit patroon van een toenemende erfelijkheid werd ook gevonden in het onderzoek van Bartels et al. Figuur 2 geeft de percentages verklaarde variantie door de verschillende variantie componenten voor verschillende leeftijden.

Het hier besproken model berust onder meer op de aannames dat interactie tussen genen en omgeving en correlatie tussen genen en omgeving afwezig zijn. We volstaan hier met de mededeling dat de aannames die aan het model ten grondslag liggen expliciet getoetst kunnen worden. De geïnteresseerde lezer verwijzen wij graag

naar het boek van Neale & Cardon (1992) dat diepgaand ingaat op de verschillende manieren waarop dat kan gebeuren en hoe het model verder kan worden uitgebreid.

Een nieuwe uitdaging

De bevinding dat er erfelijke factoren ten grondslag liggen aan overeenkomsten en verschillen tussen mensen is al een interessant gegeven op zichzelf. Het kan gebruikt worden om gevonden fenotypische relaties tussen variabelen nader te verklaren. Het kan tevens van groot belang zijn voor patiënten en hun familie om te weten dat de oorzaken van een bepaalde aandoening eerder gezocht moet worden in de omgeving dan in de genen. Winst valt zeer zeker ook te behalen in de zoektocht naar specifieke genen. Immers, als er genetische factoren ten grondslag liggen aan overeenkomsten en verschillen tussen mensen, dan moet het ook mogelijk zijn om deze genen te identificeren. De aandacht verschuift dan ook steeds meer naar zogenaamde koppelings- en associatiestudies (Sham, 1998). Met dergelijke studies worden de specifieke genen geïdentificeerd die verantwoordelijk zijn voor de erfelijkheid van een fenotype. *Structural equation modeling* is flexibel genoeg om het model dat we in dit stuk bespreken uit te breiden met componenten die koppeling en associatie mogelijk maken. Hiertoe kan in het padmodel bijvoorbeeld een extra latente variabele worden opgenomen die het gemeten genotype (een stukje DNA) indiceert.

Voor complexe tweelingmodellen en uitbreidingen naar koppeling en associatie zijn grote databestanden nodig waarin informatie is opgenomen van tweelingen en hun families van zowel hun fenotypen als hun genotypen. Het Nederlands Tweeling Register van de Afdeling Biologische Psychologie van de Vrije Universiteit is een voorbeeld van een dergelijk databestand⁴.

NOTEN

1. Een belangrijke schakel binnen het tweelingdesign vormt de vaststelling van de zygositeit van de tweeling. Het meest betrouwbare resultaat hieromtrent geeft de nog steeds kostbare vergelijking van het DNA. Onderzoek heeft echter uitgewezen dat een classificatie op basis van uiterlijke kenmerken en hoe vaak de kinderen worden verward door familie in meer dan 90% van de gevallen tot een correcte beslissing leidt omtrent de zygositeit.
2. Het menselijk DNA is voor een groot deel hetzelfde. Variatie tussen mensen kan alleen worden veroorzaakt door de zogenaamde polymorfe genen.
3. Indien een ADE model was gebruikt van zou de variantie gelijk zijn aan: $a^2 + d^2 + e^2$; de covariantie bij MZ gelijk aan: $a^2 + d^2$; en bij DZ gelijk aan: $a^2 + 0.25d^2$.
4. <http://www.tweelingenregister.org/>

LITERATUUR

- Bartels, M., Rietveld, M.J.H., Baal van, G.C.M., Boomsma, D.I. (2002). Heritability of Educational Achievement in 12-year-olds and the Overlap with Cognitive Ability. *Twin Research*, 5, 544-553.
- Heston, L.L. (1966). Psychiatric disorders in the foster home reared children of schizophrenic mothers. *British journal of Psychiatry*, 112, 819-825.
- Neale, M.C. & Cardon, L.R. (1992). *Methodology for genetic studies of twins and families*. Dordrecht: Kluwer Academic Publishers.
- Sham, P.C. (1998). *Statistics in human genetics*. London: Arnold Publishers.
- Theis, S. (1924). *How foster children turn out*. New York: SCAA.
- Wright, S. (1921). Correlation and causation. Part 1: Method of path coefficients. *Journal of agricultural research*, 20, 557-585.

REINOUW STOEL heeft als postdoc bij de Afdeling Biologische Psychologie van de Vrije Universiteit Amsterdam gewerkt, alwaar hij zich bezig hield met structural equation modeling van longitudinale tweelingdata op het NWO project 904-61.193 'Resolving cause and effect in the association between regular exercise and psychological well-being'. Momenteel werkt hij als Universitair Docent Methodenleer en Statistiek bij de Afdeling Pedagogiek en Onderwijskunde van de Universiteit van Amsterdam. E-mail <rstoel@fmg.uva.nl>.

DORRET BOOMSMA is hoogleraar gedragsgenetica bij de Afdeling Biologische Psychologie van de Vrije Universiteit Amsterdam. E-mail <di.boomsma@psy.vu.nl>.