

Drie generaties: Het XYZ model

Harry Ganzeboom, Ayşe Güveli en Ruud Luijkx¹

1. Introductie

In deze bijdrage presenteren we een methode van modellering van sociale reproductie en sociale mobiliteit tussen drie generaties binnen een familie. De hamvraag in deze situatie is of er sprake is van directe effecten van grootouders op kleinkinderen, dan wel dat sociale reproductie via een ketenmodel verloopt waarin grootouders hun kleinkinderen alleen maar beïnvloeden via de tussenliggende generatie, voor zover ze die hebben beïnvloed. Het onderzoek naar deze vraag wordt belemmerd door de versturende invloed van (random) meetfouten: uit de analyse van elementaire causale modellen met meetfouten (Kelley en Featherman, 1973a, 1973b) is bekend dat de meetkwaliteit van mediërende variabelen de geschatte verhouding tussen directe en indirecte effecten beïnvloedt. Voor meetfouten in de eerste en derde generatie geldt dat niet: zij verzwakken weliswaar het totale (=directe + indirecte) effect van eerste op derde generatie, maar veranderen niet de verhouding tussen direct en indirect en hebben derhalve geen invloed op de vraag of er een direct (grootouder) effect bestaat. In de praktijk van het drie-generaties onderzoek zijn meetfouten echter moeilijk op te sporen en te corrigeren: gegevens worden vaak verzameld via *proxy* ondervraging en hebben betrekking op dichterbij en verder weg liggende episodes met alle mogelijke gevolgen voor vertekeningen.

Als alternatief voor observatie van de drie generaties is wel voorgesteld om gegevens over familienetwerken alleen in de derde generatie te verzamelen, en de effecten van de eerste generatie af te leiden uit de gelijkens tussen *siblings* (broers en zussen), respectievelijk *cousins* (neven en nichten), waarbij de ouders (ooms en tantes) en de grootouder latent blijven. Het was voor ons

¹ Harry Ganzeboom is hoogleraar Methoden en technieken van sociaal-wetenschappelijk onderzoek aan de Vrije Universiteit Amsterdam. Ayşe Güveli is een promovendus van Wout en werkt momenteel als fellow aan de Universiteit van Essex. Ruud Luijkx is universitair hoofddocent aan de Universiteit van Tilburg en gepromoveerd bij Ultee.

een open vraag of in dergelijke familiedata het probleem van meetfouten niet speelt: siblingmodellen brengen totale familie-effecten in beeld, maar corrigeert dat ook de meetfout?

In de praktijk van het drie-generaties onderzoek zijn er nog twee lastige problemen: observaties van multi-generatie familienetwerken zijn welhaast natuurnoodzakelijk incompleet en ook indien men veronderstelt dat de wel waargenomen eenheden een *random* steekproef uit het oorspronkelijke netwerk vormen, zijn deze eenheden van elkaar afhankelijk binnen een ingewikkelde *multi-level* structuur. Het tegelijk rekening houden met incomplete en afhankelijke waarnemingen is een statistische uitdaging.

In deze bijdrage pakken we deze problemen in eerste instantie aan via gesimuleerde data: we genereren complete en gebalanceerde familie-netwerken met realistische ouder-kind associaties, onder twee scenario's: (A) het ontbreken van enig direct grootoudereffect, (B) het wel optreden van een aanzienlijk direct grootoudereffect. Vervolgens worden deze gesimuleerde data vervuld met *random* meetfouten en analyseren we in hoeverre verschillende analyse-modellen de initiële parameters terugvinden. Onze conclusie is dat het hier voorgestelde XYZ model (geschat via *simultaneous equations [SEM]*) de juiste parameters (inclusief de aangebrachte meetfout) terugvindt bij een compleet geobserveerd netwerk. Een verrassing voor ons was dat men ook de meetfout en de causale parameters correct kan terugvinden in diverse incompleet geobserveerde netwerken, ook zonder dat sprake is herhaalde meting binnen personen. Dat had geen verrassing behoeven te zijn als we ons gerealiseerd hadden dat het observeren van meerdere afstammelingen van een persoon nu eenmaal een herhaalde observatie van die betreffende persoon is. De zaak ligt echter complexer wanneer we alleen uitgaan van gegevens van de derde generatie en de tweede en eerste generatie latent laten. Het is ons niet gelukt in dat geval via een SEM model de oorspronkelijke parameters terug te schatten, met name niet wanneer de data zijn gegenereerd met een gesimuleerde meetfout.

Ten slotte presenteren we een empirische toepassing van het door ons voorgestelde XYZ model. De overdracht van islam-religiositeit onder drie generaties van Turken blijkt zeer sterk onderhevig te zijn aan directe grootouder-effecten.

uitgeoefend op verschillende generaties van sociale mobiliteitsonderzoekers (Glass, 1954; Duncan, 1966; Biblarz et al., 1996; Erola en Moisisio, 2007). In de standaard opzet van het sociaal mobiliteitsonderzoek analyseren we de gelijkenis tussen twee generaties, van ouders en kinderen. Maar ouders hebben ook ouders, en kinderen krijgen ook weer kinderen. Is de gelijkenis tussen grootouders en kleinkinderen nu een eenvoudige rekensom (=product) van de twee betrokken intergenerationele effecten, of is er meer aan de hand, doordat grootouders hun kleinkinderen ook beïnvloeden buiten de tussenliggende generatie om? De veronderstelling dat er substantiële directe grootoudereffecten bestaan, duikt daarbij in verschillende varianten op. Algemeen is de veronderstelling dat grootouders erop uit kunnen zijn voor hen negatieve uitkomsten te corrigeren: als grootouders bijvoorbeeld zien dat in de generatie van hun kinderen sociale daling is opgetreden, kunnen ze extra in hun kleinkinderen investeren om op lange termijn de maatschappelijke positie van de familie veilig te stellen. Hoe dit zou werken met grootouders die zelf een lagere status dan hun eigen kinderen hebben (een vaker voorkomend geval), is echter niet zo duidelijk. Een tweede vaak aangevoerde veronderstelling is het vaak voorkomt dat kleinkinderen meer door hun grootouders worden opgevoed dan door de tussenliggende generatie. Dat kan samenhangen met het samenlevingstype (in modern Nederland lijkt dit bijvoorbeeld een gegroeide invulling van het grootouderschap, daar waar in het verleden grootouders te vroeg waren overleden of te versleten om nog veel te kunnen doen), en kan bijvoorbeeld gestimuleerd worden door full-time arbeidsparticipatie van beide ouders of feitelijke afwezigheid van (een van) beide ouders door overlijden, ziekte of migratie. We zullen op de verschillende mogelijke achtergronden van grootoudereffecten hier niet dieper ingaan, maar constateren slechts dat in al deze gevallen de feitelijke sociale reproductie wordt onderschat wanneer men alleen op de uitkomsten van het gebruikelijke twee-generatie onderzoek afgaat.

3. Simulatie

We analyseren in eerste instantie zeer eenvoudige drie-generatie data, die we zelf hebben aangemaakt. In scenario A gaan we uit van een eenvoudig ketenmodel, waarin grootouder X alleen de middelste generatie Y beïnvloedt,

en de middelste generatie Y weer het kleinkind Z: $X \rightarrow Y \rightarrow Z$. We houden het eenvoudig door alle variabelen te standaardiseren en de intergenerationele effecten aan elkaar gelijk te stellen, namelijk op 0.71. Eenvoudige toepassing van padanalyse leert dan dat grootouders en kleinkinderen 0.50 gecorreleerd zijn en dat is natuurlijk ook precies de verklaarde variantie in de tweede en de derde generatie. In scenario B is de relatie tussen eerste en tweede generatie gelijk aan scenario A, maar wordt variatie in de derde generatie gelijkelijk bepaald door middelste en grootouder-generatie: $X \rightarrow Y$, $X+Y \rightarrow Z$. Vervolgens brengen we random meetfout aan in de gesimuleerde data. Eenvoudigheidshalve leggen we op dat deze meetfout bij alle variabelen even groot is en veronderstellen een meetcoëfficiënt van 0.90 – hetgeen overeenkomt met een betrouwbaarheid van 0.81 van elk van de kenmerken.

Tabel 1 en Tabel 2 geven een overzicht van de correlaties in de onder de twee scenario's gegenereerde data, met en zonder meetfoutvervuiling, alsmede de bijbehorende regressie-analyses. De eerste resultaten van een regressie-analyse in Tabel 1c (geen meetfoutvervuiling) zijn weinig verrassend, omdat ze in feite een herberekening zijn van de manier waarop de data gegenereerd zijn. Zonder de meetfoutvervuiling zien we de parameters waarmee de gesimuleerde data gegenereerd zijn gewoon weer terug – de geringe afwijking tussen scenario en schatting komt door de randomisatie en is te interpreteren als steekproevenfluctuatie. De resultaten voor de data met meetfoutvervuiling laten zien, dat deze random meetfouten in een ketenmodel directe effecten in het leven roepen. Herhalen we deze analyse in Tabel 2c voor data gegenereerd met een direct effect (scenario B), dan zien we een soortgelijk resultaat. Zonder meetfoutvervuiling vindt een regressie-analyse de parameters terug waarmee de data gegenereerd zijn, maar met meetfoutvervuiling wordt het directe effect te hoog ingeschat: zonder meetfout is het directe effect 58% van het totale effect, met meetfout 64%.

Drie generaties

Tabel 1a: Correlaties in gesimuleerde data (N=9000 fictieve families), ketenmodel, geen meetfout

	X0	Y1	Y2	Z11	Z12	Z21	Z22
X0	1,000	,708	,709	,498	,502	,503	,497
Y1	,708	1,000	,506	,711	,708	,356	,355
Y2	,709	,506	1,000	,356	,369	,719	,707
Z11	,498	,711	,356	1,000	,500	,258	,248
Z12	,502	,708	,369	,500	1,000	,253	,257
Z21	,503	,356	,719	,258	,253	1,000	,508
Z22	,497	,355	,707	,248	,257	,508	1,000

Tabel 1b: Correlaties in gesimuleerde data (N=9000 fictieve families), ketenmodel, met meetfout

	X0	Y1	Y2	Z11	Z12	Z21	Z22
X0	1,000	,570	,572	,405	,407	,409	,405
Y1	,570	1,000	,416	,574	,577	,293	,294
Y2	,572	,416	1,000	,298	,298	,587	,564
Z11	,405	,574	,298	1,000	,406	,208	,207
Z12	,407	,577	,298	,406	1,000	,193	,211
Z21	,409	,293	,587	,208	,193	1,000	,413
Z22	,405	,294	,564	,207	,211	,413	1,000

Tabel 1c: Regressie-analyse van Z-variantie, met ketenmodel gesimuleerde data

	geen meetfout		met meetfout	
X	0,498	-0,010	0,405	0,116
Y		0,718		0,507
R2	24,8%	50,5%	16,4%	33,8%

Tabel 2a: Correlaties in gesimuleerde data (N=9000 fictieve families), met grootoudereffect, geen meetfout

	X0	Y1	Y2	Z11	Z12	Z21	Z22
X0	1,000	,708	,709	,812	,817	,812	,814
Y1	,708	1,000	,506	,810	,812	,575	,582
Y2	,709	,506	1,000	,574	,577	,817	,819
Z11	,812	,810	,574	1,000	,775	,659	,666
Z12	,817	,812	,577	,775	1,000	,660	,667
Z21	,812	,575	,817	,659	,660	1,000	,775
Z22	,814	,582	,819	,666	,667	,775	1,000

Tabel 2b: Correlaties in gesimuleerde data (N=9000 fictieve families), met grootouder-effect, met meetfout

	X0	Y1	Y2	Z11	Z12	Z21	Z22
X0	1,000	,570	,572	,651	,656	,655	,656
Y1	,570	1,000	,416	,649	,656	,462	,462
Y2	,572	,416	1,000	,461	,467	,658	,657
Z11	,651	,649	,461	1,000	,628	,526	,530
Z12	,656	,656	,467	,628	1,000	,534	,529
Z21	,655	,462	,658	,526	,534	1,000	,618
Z22	,656	,462	,657	,530	,529	,618	1,000

Tabel 2c: Regressie-analyse van Z-variantie, met grootouder-effect gesimuleerde data

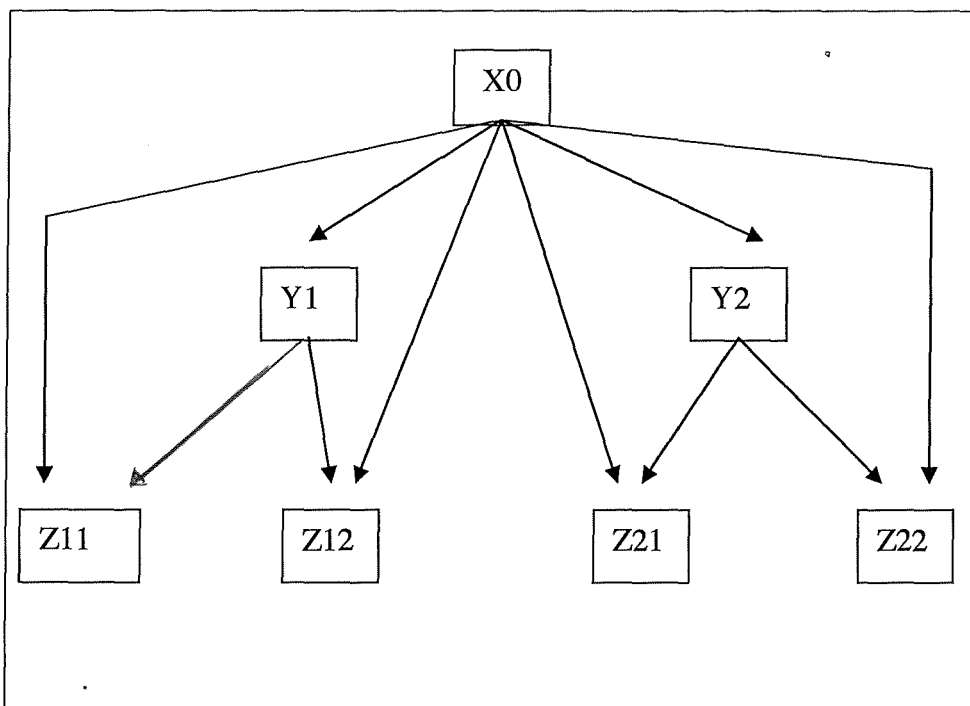
	geen meetfout		met meetfout	
X	0,812	0,478	0,651	0,416
Y		0,472		0,413
R2	65,9%	77,1%	42,3%	53,8%

In gesimuleerde data kennen we de meetfoutvervuiling. In echte data is dat natuurlijk niet het geval en moeten we hiervan een schatting maken. De meest aangewezen methode hiervoor is herhaalde meting van de betreffende indicatoren, maar deze mogelijkheid is door praktische problemen vaak moeilijk te realiseren. Wij stellen als alternatief voor deze herhaalde meting te ontleen aan het toevoegen van te meten personen in het familienetwerk. We gaan ervan uit dat elke grootouder X twee kinderen Y1 en Y2 heeft, en elke van deze twee kinderen (de middelste generatie) ook weer twee kinderen (met de symbolische namen Z11 Z12 Z21 22) heeft (zie Figuur 1). Tezamen creëren we dus zeven personen in het familienetwerk, waarin onder scenario A – zonder meetfout – *siblings* 0.50 met elkaar correleren en *cousins* 0.25 (zie Tabel 1a). Onder scenario B – met meetfout – is dat 0.41, respectievelijk 0.20 (zie Tabel 2b). Vanuit de grootouder gedacht lopen er vier familielijnen (*lineages*) en kunnen we een eventueel direct grootouder-effect dus vier maal te zien krijgen. De kleinkinderen hierin zijn welsiwaar vier verschillende personen, maar ze hebben een enkele grootouder gemeen, en vallen uiteen in twee familietakken die een middelste generatie gemeen hebben. Men zou zo'n structuur kunnen analyseren in termen van een geneste of hiërarchische *multilevel* structuur, maar dit schenkt geen aandacht aan het feit dat op elk

Drie generaties

niveau hetzelfde kenmerk wordt gemeten en dat dit met meetonbetrouwbaarheid gebeurt.

Figuur 1: Het volledige XYZ-model



Wanneer we het familienetwerk opvatten als een causaal model voor een covariantie-matrix dat geschat kan worden via *simultaneous equations model* [SEM] techniek, zoals LISREL, dan valt onmiddellijk op dat er eigenlijk slechts een zeer beperkt aantal parameters te schatten is. Onder scenario A zijn er twee causale parameters te schatten, onder scenario B drie. We beschikken in de volledige (gestandaardiseerde) data over $(49-7)/2=21$ vrijheidsgraden en dat opent de mogelijkheid om ook de meetbetrouwbaarheid te schatten uit het model. Wanneer we veronderstellen dat deze meetbetrouwbaarheid voor alle zeven personen hetzelfde is, gaat het om een enkele extra parameter. De afhankelijkheid tussen de familieleden binnen de familielijnen en -takken wordt in deze aanpak gemodelleerd doordat alle kenmerken als eigenschappen

van de werkelijk onafhankelijk waargenomen eenheden (de families zelf) worden opgevat.

We kunnen op dit basisdesign variëren, bijvoorbeeld door de veronderstelling los te laten dat bij alle generaties met dezelfde betrouwbaarheid gemeten is. Maar we kunnen ook veronderstellen dat onderdelen van het netwerk ontbreken, en dat bv. drie familielijnen i.p.v. vier zijn geobserveerd, de middengeneratie ontbreekt, of de grootoudergeneratie ontbreekt. Als we niet geïnteresseerd zouden zijn in meetbetrouwbaarheid en de gevolgen daarvan, zouden al deze varianten geïdentificeerd zijn, maar wij willen juist onderzoeken wanneer de meetbetrouwbaarheid wel geschat kan worden. We kunnen bij SEM modellen gebruik maken van Full Information Maximum Likelihood (FIML) (Allison 2001) om het ontbreken van waarnemingen tegemoet te treden. FIML kan hierbij worden gezien als genuanceerde uitwerking van het bekende *pairwise deletion of missing cases* oftewel *available case analysis* (Allison, 2001). Bij FIML modelleren we een paarsgewijs samengestelde correlatiematrix via een SEM model, maar worden de schattingen van de parameters gewogen met de N van de betrokken correlaties. Berust de schatting van een bepaald effect op correlaties met grote N, dan krijgt het een kleine SE, terwijl een effect dat correlaties met een lage N modelleert, een hogere SE meekrijgt. Deze techniek veronderstelt dat de missing values *at random* ontstaan zijn (meestal hebben we hiervoor niet veel alternatief), maar neemt mee dat daardoor soms meer en soms minder *cases* ter beschikking staan.

Tabel 3 geeft een overzicht of in een aantal verschillende varianten van mogelijke onderzoeksdesigns de meetbetrouwbaarheid is geïdentificeerd. We beperken deze oefening tot de data die gesimuleerd zijn onder scenario B (met het directe grootouder-effect) en rapporteren in welke gevallen de vier effecten, inclusief de meetbetrouwbaarheid, geïdentificeerd zijn. Het blijkt dat als we in een tak één of beide kleinkinderen niet observeren, het model prima schatbaar blijft. Als we in beide takken maar één kleinkind observeren, is het ook goed. Een voor ons verrassend resultaat is dat het model ook schatbaar blijft als je maar één tak volledig observeert, dat wil zeggen een enkele middengeneratie met twee bijbehorende kleinkinderen. Maar het werkt niet als we data hebben over twee leden van de middengeneratie en slechts een enkel kleinkind. Wel gaat het weer goed wanneer we in beide takken een enkel kleinkind observeren.

Drie generaties

Tabel 3: Identificatie van het model (met betrouwbaarheidscoëfficiënt) in verschillende soorten incomplete gesimuleerde data

Design	Data	NVAR	NDF(0)	Identificatie	NDF
1	Compleet, X0 Y1 Y2 Z11 Z12 Z21 Z22	7	21	OK	17
2	3 familielijnen: X0 Y1 Y2 Z11 Z12 Z21	6	15	OK	11
3	Alleen in 1 familietak kleinkinderen: X0 Y1 Y2 Z11 Z12	5	10	OK	6
4	Alleen 1 familietak, met twee kleinkinderen: X0 Y1 Z11 Z12	4	6	OK!!	2
5	Geen grootvader: Y1 Y2 Z11 Z12 Z21 Z22	6	15	Niet	-
6	Alleen maar kleinkinderen: Z11 Z12 Z21 Z22	4	6	Niet	-
7	Een enkele familielijjn + oom: X0 Y1 Y2 Z11	4	6	Niet	-
8	Twee familielijnen in twee takken: X0 Y1 Y2 Z11 Z21	5	10	OK	6

Al deze designs leiden tot vrijwel dezelfde parameterschattingen, en wel de getallen die we erin hebben gestopt bij het construeren van de gegevens. Hoewel er steeds dezelfde parameterschattingen uitkomen, betekent dit niet dat de oplossingen helemaal identiek zijn. Belangrijk verschil zit in het onderscheidend vermogen oftewel de geschatte steekproeffluctuatie van de parameters, als uitgedrukt in de *sampling errors* (SE) of t-waarden (Tabel 4). Het verlies in onderscheidend vermogen is betrekkelijk gering in design 3, waarvoor we gegevens onder vijf familieleden moeten verzamelen. Verzamelen we dezelfde gegevens onder vijf familieleden volgens design 8, dan is het verlies groter. Je kunt dus beter twee *sibling* kleinkinderen dan twee *cousin* kleinkinderen in je design hebben. Het enige design met vier familieleden dat geïdentificeerd is, levert het grootste verlies aan scherpte op, en het meest wat betreft de overdracht van de eerste op de tweede generatie. Dat was ook wel te verwachten, want dit is het enige design met slechts één vertegenwoordiger van de middengeneratie.

Tabel 4: Onderscheidend vermogen van verschillende geïdentificeerde modellen, afgemeten aan t-waarden, gesimuleerde data, N=9000

Parameter	Waarde	Design				
		1	2	3	4	8
Betrouwbaarheid λ	0.90	103.7	101.4	98.2	86.5	96.4
X \rightarrow Z	0.47	56.5	50.1	41.1	37.5	45.8
Y \rightarrow Z	0.47	56.0	49.6	41.5	37.6	39.7
X \rightarrow Y	0.71	81.1	80.4	79.2	47.4	73.2

In empirisch onderzoek zullen we slechts zelden een volledige observatie van het familienetwerk kunnen inzetten. Een veel gebruikt design is dat de gegevens van grootouders, een enkel lid van de tweede generatie en diens kind te zien zijn. Zulke gegevens ontstaan bijvoorbeeld als je een respondent niet alleen vraagt over zichzelf, maar ook over diens ouders en grootouders. In zulke gegevens kun je de betrouwbaarheid niet via het XYZ-model schatten. Maar dan kan wel als je dezelfde persoon ook vraagt naar kenmerken van zijn of haar sibling! Maar het werkt weer niet als je in plaats daarvan dezelfde persoon vraagt naar de kenmerken van een oom/tante!!

4. Toepassing: Overdracht van islam-religiositeit in Turkse families

Een praktische toepassing van het XYZ-model kunnen we vinden in gegevens van het LineUp project (Güveli et al. 2008), waarvoor in de zomer van 2010 *pilot* data zijn verzameld in het district Şarkışla, Turkije. Geïnspireerd door de Nijmeegse Familie-enquêtes worden in dit project data in familie-netwerken verzameld, beginnend bij Turkse mannen die in de jaren '60 gemigreerd zijn naar West-Europa of dit hadden kunnen doen (de controlegroep), en daarna bij twee van hun kinderen, en voor elk van deze twee kinderen (de middengeneratie) bij weer twee kinderen, maximaal dus vier kleinkinderen van de oorspronkelijke voorvader. Het behoeft geen betoog dat het verzamelen van dit soort data een zeer moeizame zaak is, zeker waar we in dit geval niet volstaan met *proxy* berichtgeving, maar alle betrokkenen (voorzover bestaand en nog in leven) persoonlijk ondervraagd hebben.

In het pilotproject over Şarkışla zijn tot op heden gegevens van 785 respondenten verzameld, die tezamen deel uitmaken van 263 families. Van 56 families is echter met slechts een lid gesproken. Deze families bieden geen informatie over intergenerationele overdracht en worden verder buiten beschouwing gelaten. Van de resterende 206 families hebben we dus gemiddeld

ruim drie leden in de data. Er zijn echter slechts vier families in de data met alle zeven leden. Ook observeren we maar 100 complete familielijnen, die bij elkaar gesitueerd zijn in 54 families.

In termen van incompleetheid bieden deze gegevens een enorme uitdaging. De incompleetheid is niet gestructureerd via de hierboven aangegeven systematiek. In de LineUp data komt het bijvoorbeeld ook voor dat we alleen een kleinkind en de grootouder zien, of alleen twee kleinkinderen. Juist deze situatie leent zich bij uitstek voor modellering via FIML (*Full Information Maximum Likelihood*). Het is ons echter niet gelukt om in de empirische data een FIML schatting te laten convergeren en we hebben ons voor dit gedeelte moeten terugtrekken op traditionele pairwise deletion of missing values: we analyseren een 'available case correlation matrix' op basis van de veronderstelling dat deze betrekking heeft op 206 families.

Een van de onderwerpen van het LineUp onderzoek is de overdracht van religiositeit tussen generaties, hetgeen in deze context neerkomt op islam-religiositeit. De achterliggende vragen hebben te maken met de invloed van migratie en de invloed van generatie op religiositeit, maar daar gaat het ons om deze bijdrage niet om. Hier willen we de zaak methodologisch bekijken. Religiositeit is gemeten aan de hand van zeven indicatoren, die verwijzen naar de zonden- en deugdencatalogus van de islam: de vijf centrale plichten (geloofsgetuigenis, bidden, communaal bidden (moskeebezoek), naastenliefde), twee consumptieverboden (alcohol, varkensvlees) en een algemene houdingsvraag (zie Tabel 5). Voor vrouwen is er de vraag naar de hoofddoek. Een traditionele betrouwbaarheidsanalyse resulteert in Cronbach's α van 0.81. Deze coëfficiënt is in alle drie generaties nagenoeg hetzelfde. Meetonbetrouwbaarheid leidt tot afzwakking van de relatie van het betrokken concept met andere variabelen. De mate van afzwakking ('attenuatie') is gelijk aan $1-\sqrt{\alpha}$ (Carmines en Zeller, 1979), in ons geval dus op 0.90 geschat.

Tabel 5: Indicatoren van islam-religiositeit

	N	Min	Max	Gem.	Std. Deviatie
Hoe belangrijk is religie?	756	1	5	1,89	1,12
Frequentie moskee bezoek?	752	1	6	3,61	1,73
Frequentie bidden?	738	1	7	4,11	2,49
Frequentie vrijwilligerswerk t.b.v. religie?	739	1	6	4,36	1,42
Dragen van hoofddoek?	287	1	3	2,20	0,98
Eten van varkensvlees? (-)	739	1	3	2,80	0,59
Drinken van alcohol? (-)	764	1	3	2,48	0,82
Cronbach's alpha	0.813				
Cronbach's alpha zonder R6	0.819				

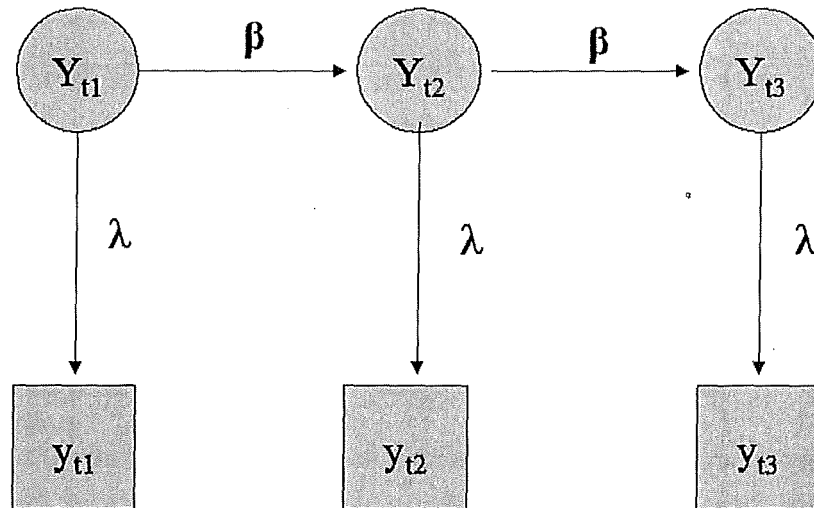
Bron: Güveli et al. 'Sarkisla pilot data'

Cronbach's α is een schatting van betrouwbaarheid op basis van herhaling van vragen binnen hetzelfde meetmoment. Deze methode wordt ook wel interne consistentieschatting genoemd. De echte betrouwbaarheid is de stabiliteit tussen metingen als de onderliggende variabele stabiel is. Het is bekend dat Cronbach's α van deze betrouwbaarheid een enigszins vertekende schatting oplevert. Cronbach's α berust op de veronderstelling van volledige homogeniteit van de samenstellende indicatoren (d.w.z. alsof het werkelijk dezelfde vraag is die meerdere keren gesteld is) en dat is nagenoeg altijd een fictie (zo ook hier), omdat voor 'herhaling' van vragen bij uitstek alternatieve inhouden worden gekozen, met een eigen unieke betekenis. Betere schattingen verkrijgen we via een *test-retest* methode, in het bijzonder via een simplex model (Alwin, 2007), waarvan ons XYZ-model een variant is. Het model achter simplex schatting van betrouwbaarheid staat afgebeeld in Figuur 2. Identificatie van de meetbetrouwbaarheid λ berust op twee restricties:

- λ is hetzelfde op de drie meetmomenten,
- Y_{t1} beïnvloedt Y_{t3} alleen via Y_{t2} , er is geen direct effect: het is een causaal ketenmodel.

Met het hierboven ontwikkelde XYZ-model kunnen we deze echte betrouwbaarheid op verwante wijze schatten, ironisch genoeg om juist wel een direct effect te kunnen bepalen!

Figuur 2: Een simplex model om betrouwbaarheid λ te schatten



Tabel 6a geeft de paarsgewijze correlaties tussen alle zeven betrokken familieleden aan. Hoewel het interessant kan zijn om in dit soort data effecten van geboortevolgorde op te sporen, vereenvoudigen we door de *lineages* (familielijnen) te randomiseren wat betreft volgorde. Een regelmatig patroon van correlaties kunnen we verkrijgen door de randomisatie meerdere keren te herhalen (we hebben het vijf keer gedaan) en vervolgens (met terugweging) op elkaar te leggen.

Het eerste dat opvalt is dat alle correlaties sterk zijn, variërend tussen 0.51 and 0.74. Het is belangrijk zich te realiseren dat deze correlaties voortkomen uit een onderzoek waarin de betrokken gegevens geheel onafhankelijk van elkaar zijn verzameld: het is heel goed mogelijk dat de grootvader zijn antwoorden gaf in augustus 2010 in een Turks bergdorp aan een Turkse interviewster, terwijl een kleindochter in Berlijn vier maanden later telefonisch antwoord gaf op dezelfde vragen in het Duits!! Een onontkoombare conclusie is daarom dat overdracht van religiositeit in deze Turkse families heel erg sterk is.

Het tweede dat opvalt is dat de correlaties allemaal betrekkelijk homogeen zijn en bijvoorbeeld niet zichtbaar verminderen wanneer we grootvader-middengeneratie vergelijken met grootvader-kleinkinderen.

Tabel 6a: Geobserveerde correlaties tussen islam-religiositeit in familienetwerken van drie generaties (N=205 families), gerandomiseerde volgorde van familielijnen

	X0	Y1	Y2	Z11	Z12	Z21	Z22
X0	1	,632	,614	,652	,652	,674	,742
	100	83	84	23	22	27	28
Y1	,632	1	,640	,685	,635	,631	,674
	83	171	141	60	60	49	48
Y2	,614	,640	1	,640	,646	,680	,737
	84	141	172	50	53	59	59
Z11	,652	,685	,640	1	,595	,595	,667
	23	60	50	66	45	25	21
Z12	,652	,635	,646	,595	1	,508	,545
	22	60	53	45	66	25	27
Z21	,674	,631	,680	,595	,508	1	,666
	27	49	59	25	25	64	44
Z22	,742	,674	,737	,667	,545	,666	1
	28	48	59	21	27	44	64

X0: Eerste generatie (grootouder)

Y1, Y2: Tweede generatie (kinderen)

Z11, Z12: Derde generatie kleinkinderen via Y1

Z21, Z22: Derde generatie kleinkinderen via Y2

Een volledig beeld verkrijgen we door de schatting van het boven ontwikkelde XYZ-model. De resultaten zijn:

$$Y1 = 0.918 * X0$$

$$Y2 = 0.918 * X0$$

$$Z11 = 0.399 * X0 + 0.559 * Y1$$

$$Z12 = 0.399 * X0 + 0.559 * Y1$$

$$Z21 = 0.399 * X0 + 0.559 * Y2$$

$$Z22 = 0.399 * X0 + 0.559 * Y2$$

Algemene meetbetrouwbaarheid λ : 0.855

Ten overvloede geven we zes regressievergelijkingen weer, maar het zal duidelijk zijn dat er eigenlijk maar twee vergelijkingen zijn met drie te schatten parameters, plus de schatting van de meetbetrouwbaarheid. De uit model model verwachte correlaties zijn bijeen gezet in Tabel 6b.

Drie generaties

Tabel 6b: Gemodelleerde correlaties tussen islam-religiositeit in familienetwerken van drie generaties (N=205 families)

	X0	Y1	Y2	Z11	Z12	Z21	Z22
X0	1,000	0,918	0,918	0,912	0,912	0,912	0,912
Y1	0,669	1,000	0,843	0,925	0,925	0,837	0,837
Y2	0,669	0,621	1,000	0,837	0,837	0,925	0,925
Z11	0,663	0,679	0,614	1,000	0,880	0,831	0,831
Z12	0,663	0,679	0,614	0,644	1,000	0,831	0,831
Z21	0,663	0,614	0,679	0,608	0,608	1,000	0,880
Z22	0,663	0,614	0,679	0,608	0,608	0,644	1,000

Boven de diagonaal: Verwachte correlaties tussen latente variabelen

Onder de diagonaal: Verwachte correlaties tussen geobserveerde variabelen

Het is nuttig om deze correlaties eens goed te bekijken en na te gaan van welke we kunnen verwachten dat ze gelijk aan elkaar zijn. Dit zijn achtereenvolgens:

- Grootouder – middengeneratie (2x) 0.918
- Grootouder – derde generatie (4x) 0.912
- Midden-generatie – eigen kinderen (4x) 0.925
- Neven/nichten-oom/tante (4x) 0.837
- Siblingcorrelatie binnen derde generatie (2x) 0.880
- Neven/nichten correlatie binnen derde generatie (4x). 0.831

We kunnen uit de geschatte parameters en de verwachte correlaties de volgende conclusies trekken:

- De meetbetrouwbaarheid λ valt wat lager uit dan we uit de Cronbach's α schatting zouden verwachten.
- In sociale reproductie correlaties doet de meetbetrouwbaarheid ($1 - \lambda$) meerdere keren mee en dat maakt correctie voor attenuatie spectaculair. Zo wordt de intergenerationale correlatie in islam-religiositeit latent op 0.92 geschat, terwijl die geobserveerd op 0.66 wordt geschat.
- Tussen grootouder en kleinkinderen is er latent een zeer sterke correlatie (0.91), en die is nagenoeg hetzelfde als tussen middengeneratie en de kleinkinderen (0.92). De mate van sociale reproductie neemt dus zeker niet af over de generaties.
- De sterke correlatie tussen middengeneratie en kleinkinderen wordt mede veroorzaakt door een sterk direct grootvadereffect. Van totale latente

correlatie tussen de eerste en derde generatie is bijna 2/3 indirect effect, maar komt meer dan 1/3 direct door de grootvader!

5. Discussie

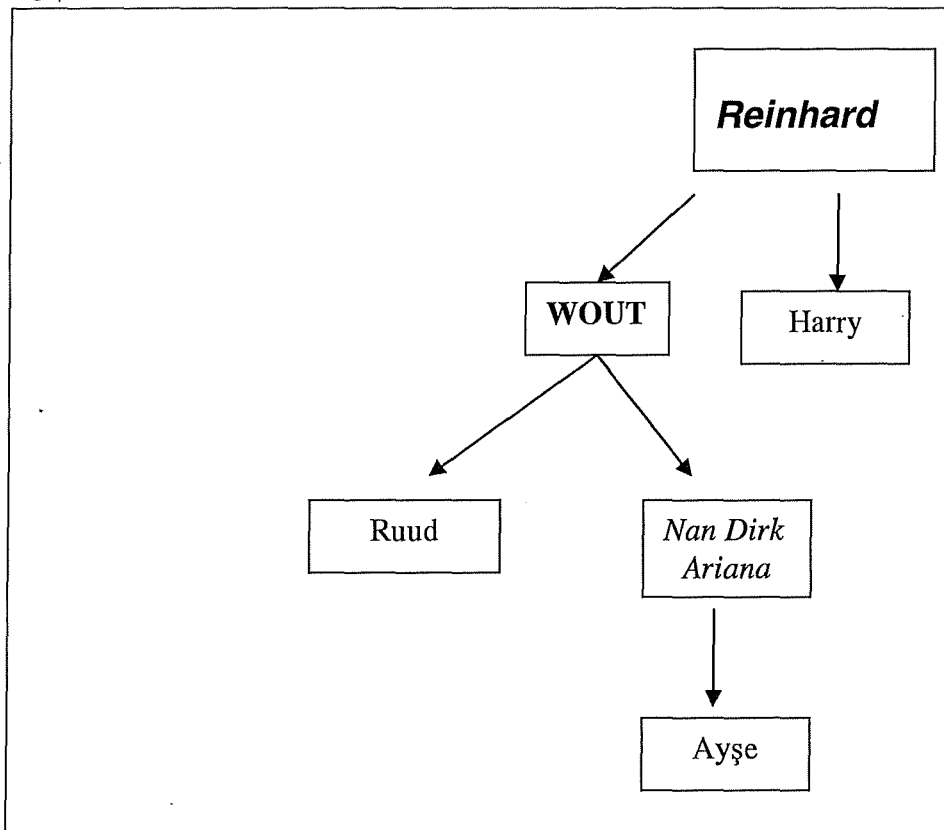
Het is niet het doel van deze bijdrage voor deze beschrijvende bevindingen een diepergaande verklaring te bieden, laat staan deze te beproeven. Als mogelijke achtergrond van de sterke overdracht stippen we aan dat sommige van de respondenten *alevi* moslims zijn, die andere opvattingen hebben over islamzonden en -deugden dan de *sunni* meerderheid. Het is denkbaar dat het sterke directe grootouder-effect berust op kleinkinderen die door hun grootouders zijn opgevoed ingeval de middengeneratie gemigreerd was. Het LineUp project, waarvoor in de loop van 2011 en 2012 de definitieve data worden verzameld op nog vier locaties in Turkije (met opvolging in West-Europa), zal een licht kunnen werpen op deze en andere verklaringen.

Onze conclusies hier zijn vooralsnog methodologische voetnoten bij zo'n toekomstige inhoudelijke analyse:

- In drie-generatie data is de hamvraag of de eerste generatie directe invloed heeft op de derde, danwel dat alle invloed indirect verloopt. Je kunt deze vraag niet beantwoorden zonder voor meetonbetrouwbaarheid te corrigeren.
- Je kunt in drie-generatie data de meetbetrouwbaarheid schatten indien je de drie generaties in ten minste twee familielijnen kunt observeren. Je moet dan wel veronderstellen dat deze meetonbetrouwbaarheid voor alle leden van het familienetwerk dezelfde is.
- Uit simulaties blijkt dat het voldoende is dat in één familie-lijn drie-generatie data worden geobserveerd, terwijl de andere lijn uit twee generaties bestaat. Het model is ook schatbaar op verschillende andere soorten incomplete familienetwerken.
- Familienetwerken van drie generaties zijn welhaast per definitie incompleet en vertonen een rijke schakering aan *missing value* patronen. Het *missing value* probleem hebben we in de gesimuleerde data goed kunnen oplossen via FIML modellering van paarsgewijs gecorrleerde data. In het concrete geval van onze LineUp data over islam-religiositeit wilde het computerprogramma LISREL vooralsnog niet convergeren en hebben we moeten volstaan met een meer traditionele *missing values* techniek.

De gepresenteerde methode lijkt ons aantrekkelijk om het hoofd te bieden aan de drie problemen waarvoor drie-generatie modellen ons stellen: meetonbetrouwbaarheid, incomplete waarnemingen en afhankelijke waarnemingen. Overdracht binnen families kan echter ingewikkelder zijn dan in het model verondersteld. Identificatie van het model berust onder meer op de veronderstelling dat ooms/tantes hun neven/nichten niet beïnvloeden anders dan vanuit de gemeenschappelijke voorvader. In Figuur 3 geven we ons over aan een moment van reflexiviteit. Berust de gelijkenis tussen Ganzeboom en Ultee uitsluitend op de invloed van Wippler? Komt de gelijkenis tussen Güveli en Luijkx door de invloed van Ultee?

Figuur 3: Een moment van zelfreflectie



**Problemen en Theorieën in Onderzoek:
Een staalkaart van de hedendaagse empirisch-
theoretische sociologie**

Gerbert Kraaykamp, Mark Levels en Ariana Need

Productie

Koninklijke van Gorcum BV, Assen

Alle rechten voorbehouden. Niets uit deze uitgave mag worden verveelvoudigd, opgeslagen in een geautomatiseerd gegevensbestand, of openbaar gemaakt in enige vorm of op enige wijze, hetzij elektronisch, mechanisch, door fotokopieën, opnamen of op welke wijze dan ook, zonder voorafgaande schriftelijke toestemming van de uitgever.

De uitgever heeft ernaar gestreefd de rechten met betrekking tot de illustraties te regelen. Wie desondanks meent rechten te kunnen doen gelden, kan zich tot de uitgever wenden.

NUR 130

ISBN 978-90-9026385-4