

EEN LONGITUDINAAL PERSPECTIEF OP INTEGRATIE: DE BOUW VAN EEN MIGRANTENPANEL

Funding door het KNAW Onderzoeksfonds, 2024-2027

Matthijs Kalmijn (NIDI), Ricarda Braukmann (DANS), Ruben van Gaalen (CBS), Frank van Tubergen (NIDI), Helga de Valk (NIDI)

In dit project wordt een nieuw 'migrantenpanel' gebouwd door gegevens uit surveys en registers op een unieke wijze te integreren. Een migrantenpanel bevat individuele gegevens over (een steekproef) van migranten (en een vergelijkingsgroep van niet-migranten) over een langere periode van hun leven (15-25 jaar). Methodisch wordt een nieuwe weg ingeslagen door voor het eerst een landelijk representatief, langlopend, en publiek toegankelijk panel over migrantengroepen te ontwikkelen. Theoretisch is het project innovatief omdat een veelheid van theorieën over de integratie van migrantengroepen op meer overtuigende wijze kan worden getoetst. Institutioneel verkent het project nieuwe manieren voor ontsluiting, archivering en disseminatie van data en wordt er samengewerkt met zowel DANS als het CBS. Met het te bouwen panel kunnen onderzoekers van het NIDI en daarbuiten, een grote impact hebben op de Nederlandse en internationale wetenschappelijke literatuur over de integratie van migranten en hun kinderen.

ONDERZOEKSDOEL 1. Harmoniseren van migrantensurveys

In Nederland bestaat een tekort aan goede longitudinale gegevens over migranten. Deze lacune geldt in de eerste plaats voor cross-sectionele vergelijkingen over de tijd. Dergelijke vergelijkingen worden wel gedaan door het CBS en SCP, maar meestal voor kenmerken die in de registers aanwezig zijn en deze zijn vaak beperkt tot sociaal-economische en demografische kenmerken. Hoe sociale en culturele dimensies van integratie veranderen over de tijd wordt mondjesmaat onderzocht. De gegevens zijn er wel. Immers, vanaf 1988 wordt er in migrantensurveys uitgebreid gevraagd naar een scala van sociale, culturele, en ook sociaal-economische kenmerken van de grotere migrantengroepen in Nederland (CBS 2022, SCP 2022). Migrantensurveys zijn om de vier jaar gehouden, gebaseerd op landelijke kanssteekproeven en omvatten zowel de eerste als de tweede generatie. Deze schat aan cross-sectionele data is onderbenut voor trendonderzoek.

Het eerste doel van dit project is om de elf migrantensurveys die vanaf 1988 zijn gehouden te harmoniseren en te integreren in één dataset met gemeenschappelijke variabelen (Tabel 1). Niet alle vragen zijn gelijk gebleven in de vragenlijsten, maar in een flink aantal gevallen en voor meerdere jaren zijn vragen (goed of redelijk goed) vergelijkbaar. Coderingen en naamgevingen dienen te worden geharmoniseerd en waar antwoordcategorieën zijn gewijzigd, dienen categorieën te worden gecombineerd zodat ze vergelijkbaar worden. De data dienen in een gemeenschappelijk bestand te worden gecombineerd met daarin het surveyjaar en surveydesignkenmerken als metavariabelen. Met behulp van deze integrale geharmoniseerde dataset kunnen vergelijkingen worden gemaakt van de bevolking als geheel over de tijd, maar kunnen ook geboortecohorten of immigratiecohorten gevolgd worden over de tijd. Cohort-volg-analyses zijn van groot belang voor vragen rondom sociaal-culturele integratie maar tot nog toe weinig gedaan.

ONDERZOEKSDOEL 2. Registerdata integreren en koppelen aan geharmoniseerde surveys

Een tweede lacune in het onderzoek is het gebrek aan longitudinale data binnen personen. In tegenstelling tot veel andere landen, is er in Nederland geen grootschalig en breed georiënteerd sociaal-wetenschappelijk panel opgezet. Hiermee wordt Nederland in de internationale academische wereld voorbijgestreefd door landen zoals Duitsland, het Verenigd Koninkrijk en Zwitserland die allemaal inmiddels langlopende panels hebben waarover in toonaangevende internationale tijdschriften wordt gepubliceerd. Het bekende LISS panel in Nederland (www.lissdata.nl) is mede als reactie hierop opgezet (Scherpenzeel 2009). Het NELLS is een panel voor twee migrantengroepen maar bevat slechts twee rondes (Tolsma et al. 2014). Doordat er geen langlopende paneldata zijn onder migranten kunnen veel sociaal-wetenschappelijke theorieën over integratie niet goed worden getoetst. Er is immers vrijwel altijd sprake van wederzijdse causale beïnvloeding en selectie-effecten. Bijvoorbeeld, de invloed van de buurt op sociaal-culturele integratie kan niet goed worden bestudeerd in een cross-sectie omdat de keuze in een bepaalde buurt te wonen mede afhankelijk is van al bestaande sociaal-culturele kenmerken. Op eenzelfde manier kan de relatie tussen de arbeidsmarktparticipatie van vrouwen en genderrol opvattingen niet goed worden bepaald zonder paneldata. Voor vragen naar de integratie van migranten zijn dergelijke complicaties eerder regel dan uitzondering.

Het tweede doel van dit project is om alle personen in de geharmoniseerde surveydata over de tijd te volgen in de registers van het CBS. De registerdata van het CBS bestaan uit een groot scala van afzonderlijke bestanden met hun eigen onderwerp en format die samen het Stelsel van Sociaal-statistische Bestanden (SSB) vormen (Bakker et al. 2014). Het doel is om deze te integreren tot een gemeenschappelijk dynamisch registerbestand met informatie over sociaal-economische, sociaal-demografische, financiële, geografische (buurt), en gezondheidskenmerken. Dit nieuw te bouwen registerbestand wordt vervolgens op individueel niveau gekoppeld aan de geharmoniseerde surveydata. Het CBS heeft de koppeling mogelijk gemaakt via *informed consent* procedures. Door de koppeling te maken kan een panel worden gebouwd voor migranten (en niet-migranten) dat de voordelen van surveys en registers combineert. Er zijn in de afgelopen jaren steeds meer studies verschenen waarin veranderingen bij migranten worden geanalyseerd vanuit de registers, maar dergelijke analyses missen essentiële onafhankelijke variabelen zoals taalvaardigheid, sociale integratie, normen en waarden, ouderlijke kenmerken, en zelfs vaak opleidingsniveau. Met het nieuwe panel kunnen binnenpersoonsveranderingen in

sociale en culturele kenmerken zelf niet worden gemeten maar kunnen zij wel dienen als variabelen die (veranderingen in) registerkenmerken verklaren.

ONDERZOEKSDOEL 3. Multi-actor perspectief aanbouwen

Een derde en laatste lacune in het bestaande onderzoek ligt in multi-actor gegevens. De sociologie en de demografie hebben hun sporen verdiend in onderzoek naar de manier waarop familieleden elkaars levensloop beïnvloeden (Kalmijn & Wittenberg 2008). Hier kan men denken aan de invloed van ouders op de levensloop van kinderen, intergenerationale steunuitwisseling, en de invloed van de partner op de arbeidsmarktloopbaan. Het bestaande migratieonderzoek in Nederland mist deze multi-actor aanpak doordat de gegevens er vaak niet zijn of niet altijd duidelijk is dat ze bestaan.

Het derde doel van dit project is om voor alle personen in het nieuwe panel de partner en de thuiswonende kinderen in de registers te volgen. De te meten kenmerken in de registers zijn hetzelfde als die voor de respondent. Voor de kinderen zal tevens de schoolloopbaan en de citoscore worden toegevoegd. Met deze multiactor data kan worden bekeken hoe de kenmerken van de respondenten invloed hebben op de levensloop van kinderen en partners.

ONDERZOEKSDOEL 4. Ontwikkelen van een disseminatiestrategie

Het vierde en laatste doel is het ontwikkelen van verschillende vormen van disseminatie. De eerste vorm is binnen de projectgroep. De oorspronkelijke surveydata en de registerdata zijn toegankelijk via Remote Access (RA) bij CBS Microdata. Het migrantenpanel wordt gebouwd binnen de RA-omgeving en kan daar gebruikt worden door de deelnemers van het project. Andere onderzoekers van het NIDI kunnen worden toegevoegd aan het project.

Omdat registerdata niet buiten de omgeving kunnen worden geëxporteerd dienen ook nieuwe manieren gevonden te worden om de data te dissemineren. Eén oplossing is het archiveren en publiek toegankelijk maken van de statistische softwarecode (STATA) waarmee de data worden gebouwd, samen met uitgebreide documentatie. Dit leidt tot de tweede manier van disseminatie, via wat men kan noemen virtuele data (vgl. Turek et al. 2021). Externe gebruikers dienen zelf toegang te krijgen tot de oorspronkelijke datafiles maar kunnen daarna met de open access code direct het nieuwe panel produceren door de code te runnen. DANS heeft ervaring met het publiek toegankelijk maken van (replicatie)code en werkt op dit terrein ook samen met relevante instituten als het Netherlands eScience Center.

Ten derde zal worden bekeken of het beschikbaar stellen van zogeheten synthetische data een publieke release van de data mogelijk maakt. Synthetische data zijn fictieve data gegenereerd vanuit de oorspronkelijke data; in deze data zijn de respondenten verdwenen maar de onderlinge samenhangen en achterliggende verdelingen van de variabelen behouden (Raghunathan 2021). Synthetische data worden gebruikt in datascience om medisch gevoelige data toegankelijk te maken. Het element van synthetische data – naast virtuele data – voorziet het project van een innovatieve disseminatie dimensie. Enige voorzichtigheid is hierbij geboden gezien de complexiteit van het te bouwen panel. Het gaat om het verkennen van de opties om survey-registerdata als synthetische data te publiceren.

DANS is betrokken bij verschillende initiatieven om gevoelige data beter beschikbaar te maken voor hergebruik en het project past in deze ontwikkeling. Zo werkt DANS met andere partners aan de ODISSEI Portal, een metadata portal waar onder meer de registerdata van het CBS vindbaar worden gemaakt. In het ODISSEI-consortium wordt ook aandacht besteed aan de mogelijkheden van synthetische data. Het CBS is een pilot gestart met synthetische data (www.cbs.nl/en-gb/about-us/innovation).

BIJDRAGE AAN KNAW DOELSTELLINGEN ONDERZOEKSFONDS

Met dit project beoogt het NIDI in samenwerking met DANS en CBS een nieuwe data infrastructuur te ontwikkelen waarmee bestaande maar gefragmenteerde data uit surveys én registers worden samengevoegd tot een uniek migrantenpanel. Nieuwe manieren van disseminatie worden ontwikkeld zodat er een langlopend dynamisch panel beschikbaar komt voor de onderzoeksgemeenschap. Unieke expertise van sociologen, demografen, en datascientists wordt geïntegreerd in een overkoepelend project.

Het project is innovatief omdat het een nieuwe weg inslaat in het datagebruik in de sociale wetenschappen. De afgelopen decennia zijn surveygegevens steeds vaker vervangen door registergegevens. De voordelen van deze transitie zijn bekend maar de nadelen worden langzamerhand ook duidelijk: een gebrek aan data over sociale, culturele, en gedragskenmerken alsmede een beperkt scala aan verklarende variabelen. Surveys blijven daarom nodig. Het huidige project probeert de voordelen van de twee soorten data te combineren en hoopt daarmee een bredere impact te hebben op de sociaal-wetenschappelijke onderzoekspraktijk in Nederland en daarbuiten.

Gezien de grote belangstelling voor longitudinaal onderzoek naar migranten zal het NIDI met de te bouwen data een voortrekkersrol kunnen vervullen voor onderzoek naar migratie in Nederland. De disseminatie wordt samen met DANS ontwikkeld en we maken gebruik van de bestaande infrastructuur om de data toegankelijk te maken. Binnen het project worden workshops georganiseerd waar het panel wordt gepresenteerd aan Nederlandse migrantieonderzoekers. In Nederland gebeurt op vrijwel alle universiteiten geavanceerd sociaal-wetenschappelijk onderzoek naar migranten. Via de workshops wordt geprobeerd deze onderzoekers warm te maken voor het panel en via hun input het panel aan te vullen

en te verbeteren.

OPZET EN MOTIVATIE PROJECTVORM

Momenteel hebben de aanvragers van het NIDI een pilot lopen waarin de koppeling van enkele migrantensurveys aan registerdata middels een geanonimiseerd RIN-nummer is gemaakt. Met deze pilot is onder meer gekeken naar de invloed van sociale en culturele integratie van migranten op retourmigratie (Özer et al. 2023) en naar inkomensontwikkelingen van migrantenvrouwen na echtscheiding (Hogendoorn & Kalmijn 2023). Beide projecten laten zien dat een koppeling mogelijk en zinvol is. De activiteiten die voor het project worden uitgevoerd worden beschreven in Figuur 1.

De bouw van het panel is complex en tijdrovend teamwerk. Tegelijk moet er niet in het luchtledige aan het panel worden gewerkt, een panel is pas bruikbaar als er inhoudelijke vragen beantwoord worden. Om die reden wordt gezocht naar een postdoc die technisch geschoold is en ook gemotiveerd is inhoudelijke papers te schrijven over de data. Er zal worden aangesloten bij bestaande projecten op het NIDI (met name inhoudelijk), DANS (met name voor datagebruik en hergebruik), en CBS (Sector Demografische en Sociaal Economische Statistieken). Op deze wijze kunnen meerdere KNAW-onderzoekers inhoudelijke publicaties schrijven op basis van de nieuwe data. Binnen het NIDI bestaat in alle themagroepen veel belangstelling voor het panel.

Het project wordt in samenwerking met het CBS uitgevoerd. Binnen het team Sociaaleconomisch Totaalbeeld (SET), binnen de sector Demografische en Sociaal Economische Statistieken (SES) van de divisie SER (Sociaal-economische en Ruimtelijke Statistieken) wordt al jaren in opdracht van het ministerie van SZW gewerkt aan een tweejaarlijkse publicatie over de stand van zaken ten aanzien van migranten (CBS 2022). Het CBS heeft groot belang bij een verbeteringslag wat betreft de niet-registerbronnen om de zachtere aspecten van integratie completer en het liefst longitudinaal in beeld te brengen. Het combineren en wegen/kalibreren van register data en survey data in één data-infrastructuur levert de juiste basis om integratieprocessen beter te begrijpen. Het CBS heeft naast inhoudelijke expertise over integratiestatistieken ook expertise over de door het CBS samengestelde administratieve bronnen (SSB; Bakker et al. 2014). Het SET heeft eveneens veel ervaring met het proces van micro-integratie (zie bijvoorbeeld de koppeling tussen de nazaten van Historische Steekproef Nederlandse Bevolking (HSN) en de Nederlandse bevolking in het SSB).

PERSONELE INZET

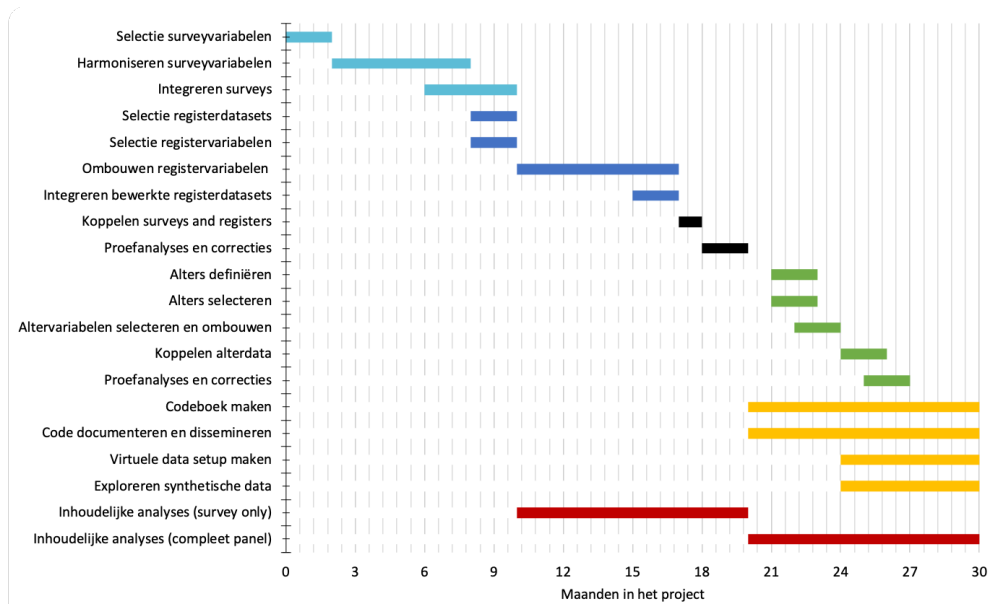
Gezien de complexiteit van het project ligt het voor de hand een postdoc te recrutereren (deze wordt aangesteld op het NIDI). De postdoc doet het methodologische werk en werkt ook aan inhoudelijke papers samen met anderen. De postdoc wordt ingebed in de NIDI themagroepen *Migrants and migration* en *Families and generations* en maakt deel uit van een projectgroep die speciaal voor dit project wordt opgericht. In de projectgroep zullen de postdoc en de vijf aanvragers zitting nemen alsmede andere junior onderzoekers van het NIDI die bij het project betrokken willen zijn.

Tabel 1. Overzicht migrantensurveys

	Naam	Jaar	Netto steekproef	Herkomstgroep									
				NL	T+M	S+A	SOM	POL	IRN	IRQ	AFG	JOE	CHI
1	SPVA	1988	5445	x	x	x							
2	SPVA	1991	9395	x	x	x							
3	SPVA	1994	8304	x	x	x							
4	SPVA	1998	14483	x	x	x							
5	SPVA	2002	8321	x	x	x							
6	SPVN	2003	5594				x					x	
7	SIM	2006*	5250	x	x	x							
8	SING	2009*	6911				x	x	x	x	x		x
9	SIM	2011*	8726	x	x	x							
10	SIM	2015*	6829	x	x	x	x	x					
11	SIM	2020	5323	x	x	x	x	x	x				

* Aanwezig in de pilot. Aanduiding herkomstgroep: NL = Nederlands (beide ouders in Nederland geboren); T+M = Turks of Marokkaans; S+A = Surinaams of Antilliaans; SOM = Somalisch; POL = Pools; IRN = Iraans; IRQ = Iraquees; AFG = Afghaans; JOE = voormalig Joegoslavisch; CHI = Chinees.

Figuur 1. Overzicht activiteiten en planning



BRONVERMELDING

- Bakker, B. F. M., Rooijen, J. V., & Toor, L. V. (2014). The system of Social Statistical Datasets of Statistics Netherlands: An integral approach to the production of register-based social statistics. *Statistical Journal of the United Nations*, 30:411-424.
- Centraal Bureau voor de Statistiek (CBS). (2022). *Integratie en samenleven 2022*. Den Haag: Staatsuitgeverij.
- Kalmijn, M., & Wittenberg, M. (red.). (2008). *Multi-actor data in survey onderzoek*. Amsterdam: Aksant.
- Özer, Ö., Remund, A., & Kalmijn, M. (2023). Back for good? Return Intentions of First-Generation Immigrants in the Netherlands. Groningen/The Hague: University of Groningen/NIDI.
- Raghunathan, T. E. (2021). Synthetic Data. *Annual Review of Statistics and Its Applications*, 8:129-140.
- Scherpenzeel, A. (2009). Start of the LISS panel: Sample and recruitment of a probability-based Internet panel. Tilburg University: Center Data.
- Sociaal en Cultureel Planbureau (SCP). (2022). *Gevestigd, maar niet thuis: Eerste bevindingen uit de Survey Integratie Migranten*. Den Haag: SCP.
- Tolsma, J., Kraaykamp, G., De Graaf, P., Kalmijn, M., & Monden, C. (2014). Design and content of the Netherlands Longitudinal Lifecourse Study: NELLS panel codebook wave 1 and wave 2. Nijmegen: Radboud University.
- Turek, K., Kalmijn, M., & Leopold, T. (2021). The Comparative Panel File: Harmonized Household Panel Surveys from Seven Countries. *European Sociological Review*, 37(3), 505-523.