



## Wat kopen Nederlanders, hoeveel bewegen ze en hoe verplaatsen ze zich?

### Smart Surveys: apps en sensoren in survey-onderzoek

Smart surveys zijn enquêtes waarbij sensoren worden ingezet voor het beantwoorden van vragen die respondenten veel tijd kosten en/of waarvoor gedetailleerde kennis en herinnering nodig is. In dit artikel bespreken we het onderzoek dat onder andere is gedaan door een Europees consortium van statistiebureaus naar de mogelijkheden en uitdagingen. We illustreren het onderzoek met toepassingen op het meten van consumptie, mobiliteit en lichaamsbeweging.

BARRY SCHOUTEN EN ANNEMIEKE LUITEN

Sommige vragen die mensen moeten beantwoorden in surveys zijn niet eenvoudig. Wie weet bijvoorbeeld precies hoeveel kilometer een verplaatsing was, of wat precies de aankomsttijd was (zoals in het CBS-verplaatsingsonderzoek gevraagd wordt), of hoeveel minuten gefietst wordt in een gemiddelde week (zoals in het CBS-gezondheids-

onderzoek). Sensoren zoals locatiesensoren en bewegingssensoren kunnen assisteren bij het beantwoorden van deze moeilijke vragen. Surveys waar sensoren worden ingezet noemen we *smart surveys*. Smart surveys kunnen het beantwoorden makkelijker en interessanter maken voor respondenten, en daarbij de kwaliteit van de

gegevens verbeteren.

Aan de andere kant, smart surveys kennen hun eigen problemen. Er kunnen nieuwe fouten worden geïntroduceerd doordat niet iedereen in staat of bereid is om sensormetingen uit te voeren. Niet iedereen heeft een smartphone, en menig respondent maakt zich zorgen over de privacy van dit soort metingen. Ook als een respondent wel bereid en in staat is tot sensormetingen, kunnen de sensoren zelf fouten maken. In veel gevallen is het dan ook nodig om de respondent te betrekken bij het interpreteren of corrigeren van de metingen. Ook is het nodig om te investeren in nieuwe logistiek, IT-infrastructuur en de ontwikkeling van (machine learning) algoritmes om de sensordata te duiden.

Sinds enkele jaren doet het CBS onderzoek naar smart surveys, waarbij we duidelijk proberen te krijgen of de potentie van smart surveys opweegt tegen de uitdagingen. De afgelopen twee jaar is dit onderzoek, met subsidie van Eurostat, opgepakt door een Europees consortium van statistiebureaus (Eurostat, 2022). De belangrijkste vragen die we moeten beantwoorden zijn:

- Waar ligt de balans tussen actieve en passieve dataverzameling; hoeveel moet de respondent doen in termen van correctie of duiding van de metingen, en wanneer wordt die inspanning voor de respondent (te) belastend.
- Welke nieuwe methodologie moet ontwikkeld worden op het gebied van machine learning en respondent benadering; wat voor strategieën moeten worden gebruikt zodat respondenten ons vertrouwen, begrijpen wat de bedoeling is, en expliciet toestemming geven om hun sensormetingen met ons te delen.
- Wat zijn de ethische en wettelijke implicaties van het verzamelen van dit soort data. Voldoet het meten met sensoren aan het principe van data minimalisatie, zoals beschreven in de wet op de privacy en gegevensbescherming (AVG/GDPR).

In de volgende paragraaf illustreren we de mogelijkheden en methodologische vragen met drie voorbeelden: het meten van consumptie, het meten van mobiliteit en het meten van fysieke bewegingen. In elk voorbeeld beschrijven we waarom het thema zich leent voor smart metingen, welke smart metingen worden gebruikt, welke nieuwe methodologie wordt ingezet op het gebied van machine learning, welke respondent-interactie gevraagd wordt, en hoe de smart survey eruit ziet.

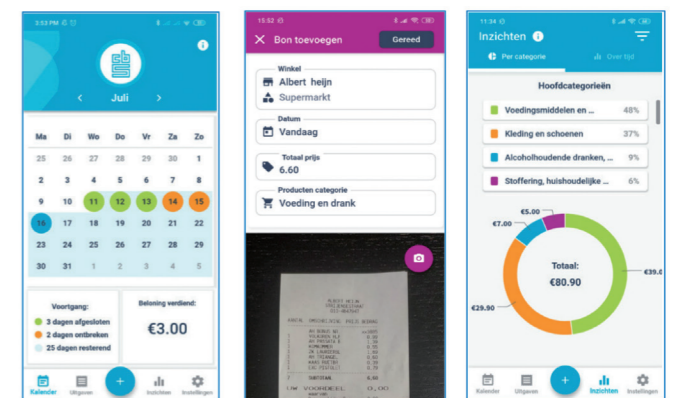
#### Toepassingen

**1. Het meten van consumptie met de HuishoudBudget App**  
Een van de meest belastende onderzoeken van het CBS is het Budget Onderzoek. Huishoudens houden een maand lang hun uitgaven van meer dan 20 euro bij, en een week lang alle uitgaven. Alle huishoudleden houden gedurende de dag een papieren dagboekje bij. Die dagboekjes worden aan het eind van de dag overgenomen in een webvragenlijst. Mede door die belasting is de respons op dit onderzoek laag. Ook is er sprake van meetfouten, doordat uitgaven vergeten worden of verkeerd worden ingevoerd.

Om de belasting voor de respondenten te verlagen en de kwaliteit van de metingen te verbeteren is een smartphone-app gebouwd waarbinnen respondenten aankopen meteen kunnen invoeren (zie figuur 1). Dat kan door aankoopbonnen in hun geheel te scannen, zodat niet langer individuele aankopen stuk voor stuk moeten worden ingevoerd. De gekochte artikelen van de scan worden via Optical Character Reading herkend. Machine learning algoritmes classificeren deze vervolgens in categorieën van COICOP (Classification of Individual Consumption according to Purpose). De COICOP is een classificatie van consumptieve uitgaven die wordt beheerd door de Verenigde Naties. De categorieën worden getoond als cumulatief overzicht in de app, zodat deelnemers een overzicht krijgen van het totaal en de aard van hun uitgaven.

In de pilots met deze app is gevarieerd met de mate van respondent-interactie door hen al dan niet de gelegenheid te geven om de resultaten van de scan te laten corrigeren.

Het laten corrigeren van de scans heeft een positieve



Figuur 1. De HuishoudBudget App

invloed op de kwaliteit van de data en lijkt ook noodzakelijk. Het percentage aankoopbedragen dat minder dan 1 euro afwijkt van het ware bedrag stijgt van 11% in de oorspronkelijke scans tot 64% in de door deelnemers gecorrigeerde scans. Ook het correcte aantal producten stijgt van 16% naar 28%. Actieve respondent-interactie lijkt voorsnog nodig om tot correcte metingen te komen in deze app. Deze voorbeelden tonen aan dat de scan zeker (nog) niet in alle gevallen leidt tot correcte identificatie en classificatie.

Details van deze pilots zijn te vinden in De Groot e.a. (2022) en Akkermans e.a. (2022).

### 2. Het meten van verplaatsingen door middel van een app

Het onderzoek 'Onderweg in Nederland' dat het CBS in opdracht van Rijkswaterstaat uitvoert, vraagt deelnemers om gedurende één dag al hun verplaatsingen van locatie naar locatie bij te houden en in te voeren in een webvragenlijst. Het onderzoek is belastend voor deelnemers: precieze tijden van verplaatsingen worden gevraagd, postcodes van begin- en eindlocatie moeten worden gezocht en ingevoerd, en van elke verplaatsing moet worden opgegeven met welk vervoermiddel deze werd uitgevoerd. Dit leidt tot meetfouten als gevolg van onzorgvuldig invullen, omdat deelnemers vertrouwen op hun geheugen, verplaatsingen vergeten of de vragenlijst invullen als de dag nog niet is afgelopen.

Smartphones bevatten locatie-sensoren die een groot deel van deze lastige taak kunnen overnemen. Door middel van GPS kan de positie van een smartphone tot 5 meter nauwkeurig worden bepaald. Dezelfde sensor kan

tonen met welke snelheid een telefoon zich verplaatst. Machine learning algoritmes bepalen wanneer iemand onderweg is en wanneer een locatie is aangedaan. Met deze informatie kan, gekoppeld aan informatie over wegenet en openbaar vervoer infrastructuur, worden afgeleid met welk vervoermiddel een verplaatsing is gedaan (figuur 2). Het is in principe mogelijk dat de deelnemer, behalve het downloaden van de app en het invullen van wat achtergrondkenmerken nog nauwelijks naar de app hoeft om te kijken. Zo ver zijn we nog niet.

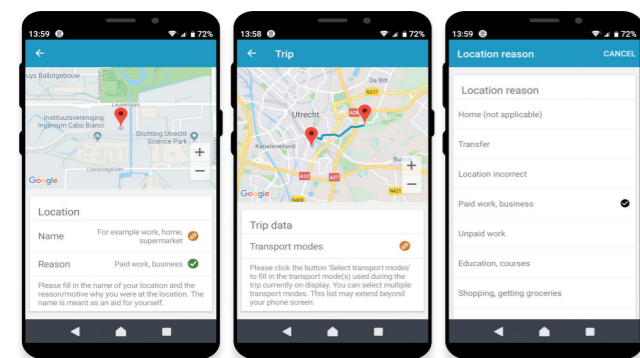
Een actieve rol voor respondenten is nog nodig om context te geven bij een verplaatsing (waarom was de respondent onderweg) en om het vervoermiddel in te voeren. Ook is de app nog niet in staat om onderscheid te maken tussen een 'echte' locatie en, bijvoorbeeld, stilstaan op de snelweg in een file. Het onderzoek naar hoeveel correcties die respondent moet geven en hoeveel het algoritme kan corrigeren is nog lopend.

Meer details over dit onderzoek zijn beschreven in McCool e.a. (2021).

### 3. Het meten van lichaamsbeweging met een bewegemeter

Waar in de vorige twee voorbeelden gebruik werd gemaakt van een smartphone app, wordt in dit voorbeeld gebruik gemaakt van externe sensors (bewegemeters) die deelnemers van het CBS krijgen (figuur 3).

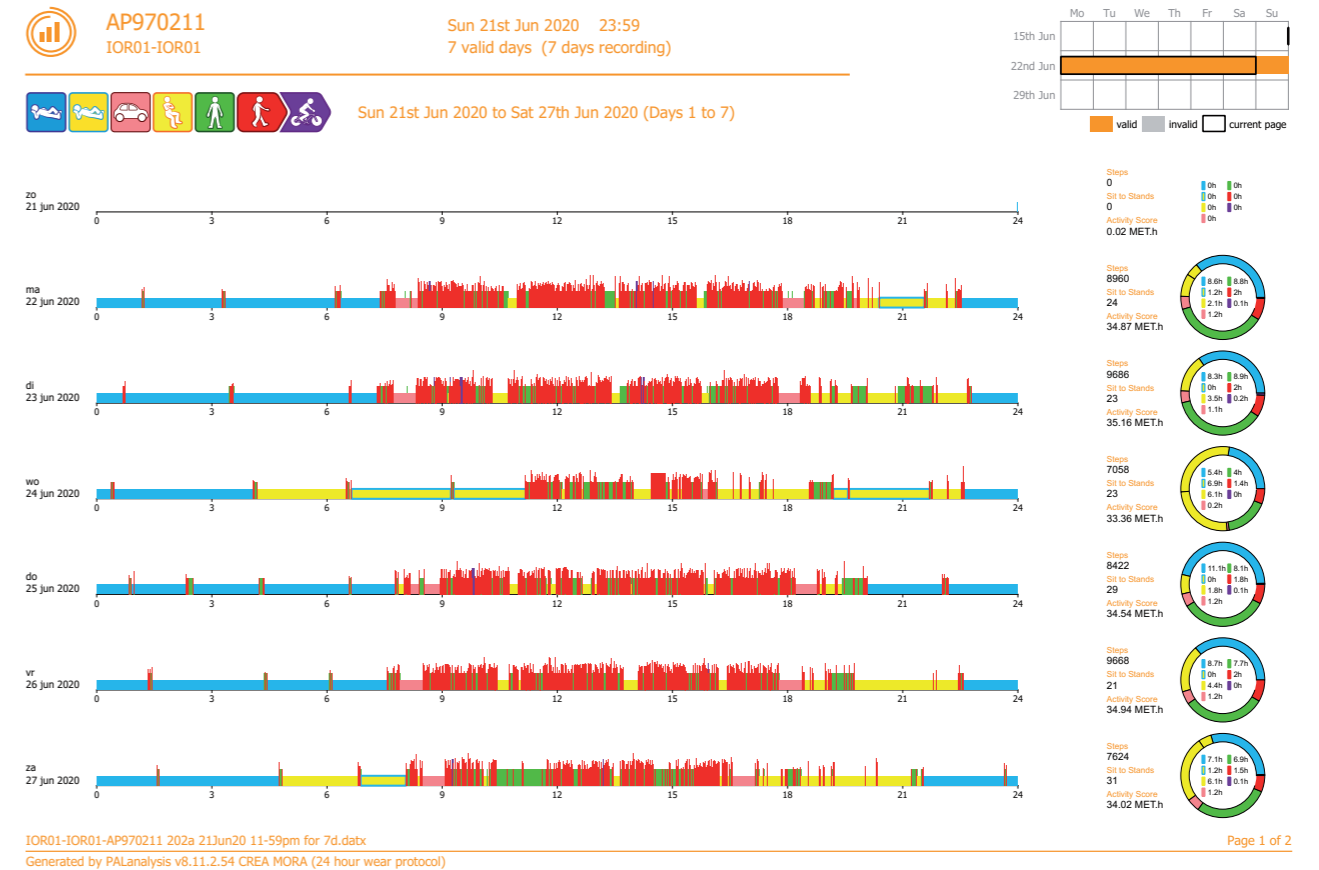
In CBS-surveys over gezondheid en leefstijl wordt lichaamsbeweging gemeten door middel van een vragenlijst. Deze vraagt respondenten om te schatten hoeveel tijd zij in een 'gemiddelde week' besteden aan sport, fietsen, huishoudelijke taken, tuinieren, etc. Dit is geen



Figuur 2. App om verplaatsingen te meten



Figuur 3. Bewegemeters



Figuur 4. Data van de bewegemeter

gemakkelijke taak, en onderzoek toont aan dat mensen de mate van lichamelijke activiteit schromelijk overschatten (bijvoorbeeld Kapteyn e.a. 2018). Het meten van lichaamsbeweging lijkt dan ook een goede kandidaat voor meting via sensoren.

In vier pilots is respondenten die een vragenlijst invulden gevraagd of zij aanvullend wilden meedoen met een onderzoek met bewegemeters. Voor het onderzoek is gebruik gemaakt van de activPAL™ bewegemeter die op het bovenbeen wordt gedragen. De meter meet de positie van het been en of er wel of niet activiteit is. Op basis daarvan berekent een algoritme of iemand ligt, zit, staat, loopt of fietst, en met welke intensiteit. De meters vergen geen interactie met de deelnemer: de data worden opgeslagen op de meter en – na terugsturen – uitgelezen en verder verwerkt (figuur 4).

In deze pilots was een belangrijke onderzoeksvraag of mensen mee willen doen met dit soort onderzoek, en wat voor soort mensen dat zijn: alleen topsporters, of ook mensen die (te) weinig bewegen. In elk van de pilots was het aandeel respondenten op de vragenlijst dat bereid was om mee te doen met aanvullend onderzoek

ongeveer 50%. Afhankelijk van de initiële respons op de vragenlijst komt dat neer op 11 tot 38% van de steekproef. De deelnemers zijn vooral mensen die voldoen aan de bewegennorm van 150 minuten minimaal matig intensief bewegen per week (RIVM, 2020). Toekomstig onderzoek richt zich op het vinden van manieren om ook minder actieve deelnemers te vinden. Zie Luiten e.a. (2022) voor een gedetailleerde beschrijving van de opzet en resultaten van de pilots.

### Discussie

In drie voorbeelden hebben we laten zien dat smart surveys kansen bieden. Mensen zijn bereid om mee te doen met dit soort metingen, hoewel die bereidheid (nog) niet op het zelfde niveau is als bij vragenlijsten. Aanvullend onderzoek is nodig hoe wij de representativiteit van de respons kunnen verbeteren.

In de inleiding is beschreven dat er drie belangrijke vragen zijn die beantwoord moeten worden om de haalbaarheid van smart surveys goed te kunnen beoor-

delen. In de voorbeelden die beschreven zijn, zijn eerste stappen gezet, maar het eindoordeel is nog verre van duidelijk en kan verschillend zijn voor verschillende toepassingen.

Methodologische vragen die er nog leven betreffen de te ontwikkelen machine learning algoritmes. Ook deze zijn feilbaar en hebben een foutmarge op de voorspellingen. Een beweging kan bijvoorbeeld met 80% kans lopen zijn en met 20% kans rennen. Bovendien zullen de algoritmes herhaaldelijk of zelfs doorlopend moeten worden getraind, bijvoorbeeld door de introductie van nieuwe producten in het Budget onderzoek. Respondenten zijn onmisbaar bij het verbeteren van nauwkeurigheid. Welke respondent-interactie noodzakelijk en niet te belastend is, hoe deze interactie op een begrijpelijke manier in te bouwen in een user interface, en hoe om te gaan met de onzekerheden in machine learning voorspellingen als respondenten het ook niet kunnen weten zijn aspecten waar nader onderzoek nodig is.

Ook met betrekking tot logistiek en IT zijn er nog legio vragen. We gebruiken privé apparaten van onze respondenten en hebben complexe IT nodig om sensoren aan te sturen. Hoe deze afdoende te beveiligen is nog onvolledig bekend.

We verzamelen nieuwe vormen van data die respondenten zelf ook niet volledig kennen. Dit roept de vraag op of respondenten wel weten waarvoor ze toestemming geven. Hoeveel en welke informatie moeten we respondenten geven opdat zij werkelijk 'informed consent' kunnen geven? De wet op de informatiebescherming roept de vraag op hoe en waar data voorbewerkt kunnen worden vanuit het gezichtspunt van data minimalisatie. Classificatie van producten in de Budget app zou bijvoorbeeld aanzienlijk nauwkeuriger en sneller kunnen als de data van het apparaat van de respondent naar het CBS gezonden konden worden, geclassificeerd, en daarna weer teruggestuurd naar de respondent. Vanwege de AVG is dit echter problematisch.

Tot slot is er nog de open vraag wat te doen als de nieuwe vormen van data niet perfect zijn, wat meestal het geval zal zijn, en hoe we moeten omgaan met meta-informatie over kwaliteit. Geldt de AVG ook voor metadata? Zo ja, hoe kunnen we dan nog weten of data wel goed

genoeg zijn, bijvoorbeeld als locatiemetingen uitbijters bevatten of gaten.

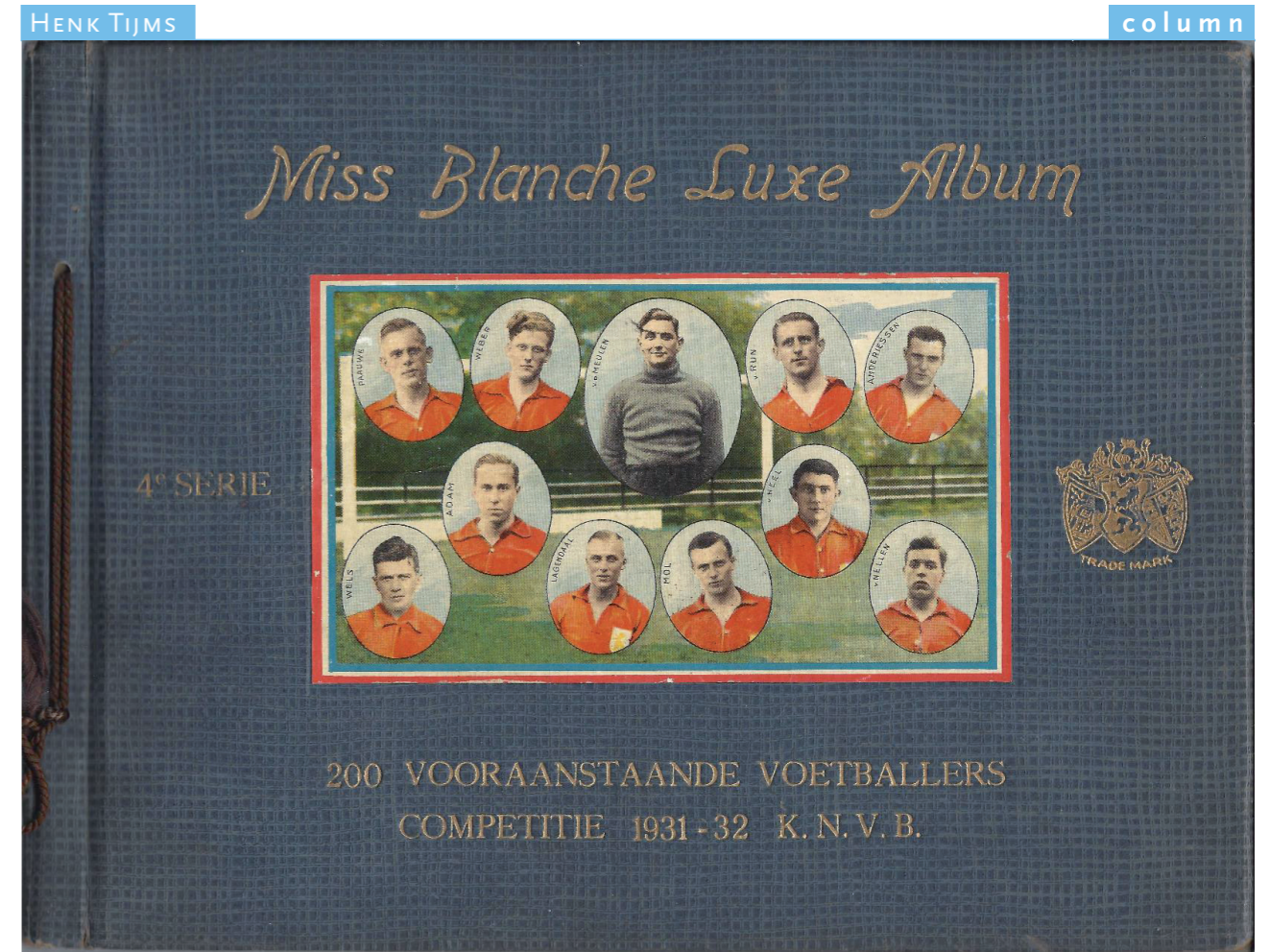
In de komende jaren is aanvullend onderzoek voorzien dat ons inzicht in smart surveys verder zal verdiepen.

#### LITERATUUR

- Akkermans, J., Rodenburg, E., Groot, J. de, Schouten, B., Martin Bernia, E., Balsa Criado, V., Horcajo, T., Gauche, C., & Osier, G. (2022). *The role of interviewers*. Geraadpleegd van CROS website ESSnet Smart Surveys.
- De Groot, J., Oerlemans, T., Rodenburg, E., Schouten, B., Lope Mariscal, A.C., Bernia, M., Poch, J., Horcajo, T., Blasé Criado, V., Gauche, C., & Osier, G. (2022). *Thematic paper on the Consumption pilot*. Geraadpleegd van [https://ec.europa.eu/eurostat/cros/system/files/220413\\_wp2\\_-\\_deliverable\\_2.1\\_-\\_consumption.pdf](https://ec.europa.eu/eurostat/cros/system/files/220413_wp2_-_deliverable_2.1_-_consumption.pdf).
- Eurostat. (2022). Projectdocumentatie ESSnet Smart Surveys, [https://ec.europa.eu/eurostat/cros/content/ess\[1\]net-smart-surveys\\_en](https://ec.europa.eu/eurostat/cros/content/ess[1]net-smart-surveys_en).
- Kapteyn, A., Banks, J., Hamer, M., Smith, J. P., Steptoe, A., Van Soest, A., & Wah, S.H. (2018). What they say and what they do: comparing physical activity across the USA, England and the Netherlands. *J Epidemiol Community Health*, 72(6), 471-476.
- Luiten, A., Toepoel, V., Schouten, B., Cierpiat-Wolan, M., Kapica, K., Szlachta, P., Lusyne, P., & Van der Beken, H. (2022). *Health – measuring physical activity*. Deliverable 2.1. Essnet Smart Statistics. Geraadpleegd van Thematic papers per pilot - Physical activity (europa.eu).
- McCool, D., Lugtig, P., Mussmann, O., & Schouten, B. (2021). An app-assisted travel survey in official statistics: Possibilities and challenges. *Journal of Official Statistics*, 37, 149-170. <http://dx.doi.org/10.2478/JOS-2021-0007>.
- RIVM. (2020). Beweegrichtlijnen. Geraadpleegd van <https://www.rijksoverheid.nl/documenten/publicaties/2020/07/07/beweegrichtlijnen>.

BARRY SCHOUTEN werkt als senior wetenschappelijk medewerker bij het Centraal Bureau voor de Statistiek in Den Haag en als bijzonder hoogleraar bij de Universiteit Utrecht. Hij is coördinator van een gezamenlijk dataverzameling-innovatienetwerk tussen het CBS en de Universiteit Utrecht. E-mail: [jg.schouten@cbs.nl](mailto:jg.schouten@cbs.nl)

ANNEMIEKE LUITEN werkt als wetenschappelijk medewerker bij het Centraal Bureau voor de Statistiek in Heerlen waar zij onderzoek doet naar de rol van sensormetingen in enquêtes. E-mail: [a.luiten@cbs.nl](mailto:a.luiten@cbs.nl)



## VOETBALPLAATJES wie heeft ze niet verzameld?

Voor het wereldkampioenschap voetbal in Qatar hebben Albert Heijn, Jumbo en andere supermarktketens klaarblijkelijk afgezien van acties met plaatjes van de voetballers die acte de présence geven op dit besmette evenement in Qatar. Het verzamelen van voetbalplaatjes is wereldwijd een rage onder de jeugd. Menig lezer van STATOR zal zich vroeger ook aan deze verslavende activiteit overgegeven hebben. Wereldberoemd zijn de voetbalalbums van de gebroeders Guiseppe en Benito Panini. In 1961 brachten ze vanuit hun simpele krantenkiosk in de Italiaanse stad Modena hun eerste album uit met plaatjes van alle voetballers die in dat jaar actief waren in de Italiaanse voetbalcompetitie Serie A. Dit was het begin van een wereldwijd imperium dat de gebroeders Panini opbouwden en dat hen schatrijk maakte. Overigens waren de Panini-albums niet de eerste in hun soort. In Ne-

derland werden jaren dertig en veertig van de vorige eeuw al diverse albums met plaatjes van sporthelden uitgebracht door het legendarische sigarettenmerk Miss Blanche van de Vittoria Egyptian Cigarette Company uit Rotterdam, een sigarettenmerk dat ook de inspiratie vormde voor de psychologische roman *Miss Blanche* uit 2004 van Nelleke Noordervliet over een Rotterdamse tabakswinkelier die zijn leven ziet veranderen door een toevallige ontmoeting met een kettingrokende Turkse schone. De Miss-Blanche-albums zijn heden ten dage ware *collectors items*, zeker het album dat in 1932 werd uitgebracht met plaatjes van tweehonderd vooraanstaande voetballers uit de eerste en tweede klas van de Nederlandse voetbalcompetitie 1931-1932, waaronder de befaamde internationals Leo Halle, Puck van Heel en Bas Pauwe. Fabrikanten van margarinemerken als Leeuwenzegel en