

•••••
Column

Lex Borghans

Anders denken

Het Sociaal Statistisch Bestand (SSB), de virtuele volkstelling van het CBS waarin informatie over de Nederlandse bevolking uit verschillende bronnen bijeen wordt gebracht, wordt langzaam realiteit voor onderzoekers. Deze data hebben een enorme potentie. Dit is echter niet vanzelfsprekend. Net zoals de eerste auto gebouwd werd als een koets, zo lijkt men nu vooral te kijken of administratieve data bestaande bronnen kunnen vervangen. Als je administratieve data echter vergelijkt met traditionele enquêtebestanden, dan heb je feitelijk veel minder informatie over veel meer personen. Vele honderden variabelen over misschien vijfduizend personen in bijvoorbeeld het Sociaal Economisch Panel (SEP) hebben plaatsgemaakt voor een handjevol variabelen over 17 miljoen mensen. Ieder statistiekboek laat echter zien dat je met informatie over vijfduizend personen vrij nauwkeurig uitspraken kunt doen over 17 miljoen mensen. Er lijkt dus sprake te zijn van een verlies.



Dit betekent in de eerste plaats dat administratieve data de traditionele enquêtes niet kunnen vervangen. Maar met administratieve bestanden – zeker in de zeer doordachte opzet van het CBS – kunnen ook dingen die met de traditionele bestanden niet konden:

Allereerst kunnen we kijken naar zeer specifieke groepen zonder dat de steekproef te klein hoeft te worden. Als we de kappers uit het SEP nemen is er bijna geen steekproef meer over. In het SSB zitten er vele tienduizenden. Specifieke groepen kunnen dus zonder statistische problemen adequaat geanalyseerd worden.

Ten tweede kunnen verbanden tussen mensen veel beter worden onderzocht. Voor iedere observatie in het bestand is ook een observatie van zijn collega, buurman en broer of zus aanwezig. Onderlinge invloeden tussen mensen kunnen dus veel beter worden geanalyseerd en vergelijkingen veel beter worden gemaakt.

Ten derde is het probleem van uitval verdwenen. Bij een panel zijn we vaak na enkele jaren al een behoorlijk aantal mensen kwijt. Met administratieve data is er alleen sprake van uitval als mensen sterven of het land verlaten. Dit betekent dat we zeker op den duur werkelijke zeer-lange-termijn-effecten kunnen gaan bestuderen.

Door op een slimme manier gebruik te maken van deze mogelijkheden kunnen belangrijke maatschappelijke vragen onderzocht worden die vroeger onbeantwoordbaar leken. Om deze nieuwe mogelijkheden werkelijk te benutten moeten we op een andere manier gaan denken over de vraag welke informatie belangrijk is: minder clustering en meer detail in classificaties, gedetailleerde informatie over relaties tussen mensen en gegevens over omstandigheden die van invloed kunnen zijn op het latere functioneren van mensen. Omdat de potentiële waarde van informatie enorm is gestegen moeten we niet te zuinig zijn, eerder meer dan minder aanvullende informatie verzamelen en wat verzameld wordt zeker niet van het toeval af laten hangen. Vooral als we meer naar langetermijnverbanden willen kijken, bepalen de keuzes die nu gemaakt worden de onderzoekinformatie die de komende decennia beschikbaar zal zijn.

Lex Borghans is hoogleraar arbeidseconomie en sociaal beleid aan de Universiteit Maastricht.

•••••