

## Opgave 3 De Big Data Paradox

### tekst 5

#### VS maakten klassieke fout bij te hoge schatting van gevaccineerden

Ze zaten er goed naast. Twee Amerikaanse onderzoeksbureaus peilden dit voorjaar de vaccinatiebereidheid onder de bevolking. Beide schatten dat ruim twee derde van de Amerikanen zijn eerste prik had opgehaald. Dat strookte niet met de cijfers waar het CDC<sup>1)</sup>, de Amerikaanse evenknie van het RIVM<sup>2)</sup>, later mee kwam. Het CDC had geturfd hoeveel vaccins waren uitgedeeld en kwam niet verder dan een vaccinatiegraad van 53 procent.

De schattingen leken zo betrouwbaar. Het ene bureau, Delphi-Facebook, had een paar keer gepeild en kreeg telkens 250.000 antwoorden binnen. Het andere, een agentschap van de overheid: 75.000 per week. En dat bijna een halfjaar lang. Beide hadden ze miljoenen reacties in hun peilbak<sup>3)</sup>.

Dat zegt dus niet alles, schrijven statistici van Harvard en Oxford in het vakblad Nature. Dit is een typisch geval van wat zij de Big Data Paradox noemen. Het is tegenwoordig heel gemakkelijk om met grote databestanden te werken. Je plukt ze uit een digitaal archief, of je schrijft, zoals in een van deze gevallen, gebruikers van Facebook aan. Maar veel is niet altijd beter, betogen de onderzoekers. (...) Terwijl de peilers zelf denken dat ze met hun grote groep een hoge mate van nauwkeurigheid hebben. En dat is in dit geval zelfs gevaarlijk. De overheid zou op basis van zo'n peiling kunnen denken dat ze de coronamaatregelen kan versoepelen. (...) De mensen op Facebook zijn wit en hoger opgeleid. Het zijn juist de zwarte, laagopgeleide Amerikanen die bij de vaccinaties verstek laten gaan. (...)

*bron: Trouw, 2021*

noot 1 CDC = Centers for Disease Control and Prevention (centra voor ziektebestrijding en -preventie)

noot 2 RIVM = Rijksinstituut voor Volksgezondheid en Milieu

noot 3 peilbak = verzameling van peilingen, oftewel metingen; dataset