

---



---

## Hoorcollege 9

Bestaande gegevens zijn vaak niet reactief. Reactiviteit kan een bedreiging zijn voor de interne validiteit.

### Bestaande data en statistieken

Bestaande data worden gebruikt voor nieuw onderzoek. Er zijn een aantal voorbeelden:

- Secundaire analyse: er zijn al gegevens alleen worden ze voor de tweede keer geanalyseerd voor een ander wetenschappelijke onderzoek. Je hebt hierbij toegang tot de oorspronkelijke data. Er is ruwe data nodig die vaak online te vinden is. Een voorbeeld waar van een site waar data gevonden kan worden is DANS voor data uit Nederland. Voor data uit de US kan de General Social Survey gebruikt worden.
- Reeds beschikbare statistieken: dit zijn gegevens die eerder verzameld zijn maar toen voor een ander doel dan wetenschappelijke analyse. Voorbeelden hiervan zijn publieke informatie van het CBS en ambtelijke statistische gegevens. Je hebt hierbij toegang tot bestaande statistieken maar niet tot de ruwe data. Veel statistieken staan online in samenvattende tabellen (voorbeeld is het CBS).

We kunnen deze methodes vergelijken. De secundaire analyse heeft de meeste flexibiliteit omdat je daarbij toegang hebt tot de ruwe data. Er zijn ook een aantal beperkingen van bestaande data/statistieken zoals missende gegevens, betrouwbaarheid en validiteit. De validiteit is bekeken voor het oorspronkelijke onderzoek en daar heb je daarna geen controle meer over. Er is dus geen controle over de kwaliteit van de operationalisatie. Een andere beperking is misplaced concreteness. Hierbij wordt heel specifiek data gegeven terwijl je niet weet of je er zo zeker van kan zijn. Een voorbeeld hiervan is Maurice de Hond die schattingen maakt over de peilingen van de politieke partijen.

### Fysieke neerslag van gedrag

Fysieke neerslag wordt in Neuman vertaald als physical evidence. Het ultieme archief is de vuilnishoop. Dit is onderzocht met het garbage project waarbij onderzoek werd gedaan naar de moderne vuilnisresten. Hiermee kan het best gekeken worden hoe mensen zich gedragen en wat ze bijvoorbeeld eten, drinken en doen. Het is non-verbaal gedrag. Het centrale idee is of we zeggen wat we doen hetzelfde is als wat we echt doen. Een belangrijke factor hiervan is sociale wenselijkheid. Er wordt dan niet reactief gemeten omdat we eerder geloven wat iemand doet dan wat iemand zegt. Er zijn ook een aantal beperkingen van onderzoek via fysieke neerslag. Een eerste nadeel is privacy, mag je wel zomaar in de spullen van anderen kijken? Dit is een ethisch aspect. Ten tweede is het vooral indirect bewijs, je weet niet de oorzaak van het vuilnis van iemand. Je kan alleen educated guess doen voor het hoe en waarom mensen wat doen. Dit heeft te maken met causaliteit. Ten derde kan je geen gevoel, mening of beleving meten van een persoon.

### Inhoudsanalyse

Een inhoudsanalyse analyseert boodschappen van de media. Ze kijken naar hoe het gecommuniceerd wordt en wat er gecommuniceerd wordt. Het bestudeert bestaande boodschappen waardoor het niet-reactief is. Het is erg geschikt voor grote hoeveelheden informatie en informatie van lang geleden.

Daarnaast is een voordeel dat het vanaf een grote afstand bekeken kan worden.

Eerst wordt het materiaal verzameld en daarna wordt een coderings-schema gemaakt. Er wordt dan vastgelegd hoe eigenschappen van boodschappen vertaald worden in variabelen. Hierdoor kan de frequentie van bepaalde boodschappen bekeken worden en bijvoorbeeld de intensiteit. Ook kan bepaald

worden welk doel je precies hebt met het onderzoek. De betrouwbaarheid kan verhoogd worden door twee onafhankelijke codeurs (onderzoekers) te gebruiken. Hierdoor kan de inter-codeur betrouwbaarheid (correlatie tussen de twee codeurs) berekend worden.

De validiteit kan verhoogd worden door manifeste en latente codes te gebruiken.

Er zijn een aantal berekeningen. Ten eerste gaat het vooral om de causaliteit, is de boodschap waar? Je hebt veel te maken met interpretaties. Want de invloed is moeilijk te bepalen.

## Systematische review en meta-analyse

Een systematische review is een systematische identificatie van de literatuur in alle relevante studies over een specifiek onderwerp. Een systematische review is een voorgeschreven en expliciete methodologie.

Een nuttige vraag is of er veel goede studies over het onderwerp bestaan. Ook bestaan er traditionele reviews. In de traditionele review is er vaak een gebrek aan regels en er zijn geen updates. Daarnaast komen veel verschillende reviewers tot verschillende conclusies. Daarom wordt er steeds meer gebruik gemaakt van systematische reviews. Deze volgen wetenschappelijke regels om fouten te kunnen reduceren. Daarnaast zijn er veel updates en wordt de methode uitgebreid beschreven zodat het gerepliceerd kan worden.

Een review is de beste manier om literatuurstudie te doen. Het is vaak voor toegepast onderzoek. Het gaat volgens een systematische methode:

- Formuleren van de reviewvraag. Deze vraag bestaat uit verschillende onderdelen. Ten eerste om wie gaat het? Vervolgens wat gaan we veranderen en hoe gaan we het vergelijken? Als laatste de vraag wat is een relevante maat om het te vergelijken? Er wordt veel gebruik gemaakt van de eigen kennis en van partijen die van belang zijn voor de vraag.
- Systematisch zoeken. Je zoekt op basis van de reviewvraag met een zoekmachine naar keuze.
- Gevonden literatuur analyseren op geschiktheid. Zijn mijn zoekcriteria goed genoeg? En de beoordeling op basis van de titel en de samenvatting. Sommige kunnen wel of niet nuttig zijn maar er zijn altijd twijfelgevallen. Deze ga je beoordelen door het hele artikel te lezen. Na het onderzoeken van de twijfelgevallen weet je of je ze wel of niet nodig hebt.
- Het beoordelen van de kwaliteit van de literatuur. Er wordt dan bijvoorbeeld beoordeeld op de ethiek, steekproeftrekking en de manier van data-verzameling.
- Samenvatten geselecteerde literatuur. Dit kan op verschillende manieren, narratief (beschrijvend), kwalitatief (coderen) en kwantitatief (meta-analyse). Een meta-analyse is het samenvatten van kwantitatieve onderzoeksresultaten. Hierdoor kan er een analyse van oorzaak en verschillen gedaan worden.

Er zijn ook een aantal nadelen. Ten eerste zijn er weinig relevante studies en iedere studie kan methodologisch mogelijk anders zijn. Als laatste is er ook nog het file-drawer probleem. Dit is het probleem van de onderzoek publicaties. Er worden te veel papers met significante resultaten gepubliceerd en minder met niet significante resultaten. Maar deze studies zijn dus ook bijna niet te vinden waardoor je het minder van twee kanten kan bekijken.

## Big data

Het zijn grote hoeveelheden gegevens. De kenmerken zijn volume, velocity (veel data is tegenwoordig online en dus makkelijk en snel te vinden) en variabiliteit (veel verschil en variatie in data). Er zijn vier soorten big data. Ten eerste de register data. Een voorbeeld hiervan is het CBS. Ten tweede transactie data, een voorbeeld is credit card data. Ten derde social media, hierbij moet bijvoorbeeld gedacht worden aan twitter en facebook. Als laatste GPS data die komt uit navigatiesystemen of bijvoorbeeld de telefoon. Er zijn een aantal beperkingen. De kwaliteit is niet altijd erg hoog. Mensen hoeven bijvoorbeeld niet altijd

de goede informatie te geven. Ook het dekkingsprobleem speelt een rol, want de steekproef van bijvoorbeeld twitter hoeft niet een representatieve steekproef te zijn voor de populatie. Daarnaast hebben ze veel te maken met privacy en de causaliteit is vaak niet te verklaren.