

Inleiding

Wat zijn paradata en welke data voor welk gebruik verzamelen?

Ann Carton

Discussiemiddag 'paradata', Nederlandstalig
Platform voor Survey-Onderzoek
Brussel, 11 maart 2010



Wat zijn paradata?

- Data
 - » Gegevens over het fenomeen dat bestudeerd wordt
- Metadata
 - » Data over de verzamelde data
- Paradata
 - » Data over het survey proces (procesdata)



Wat zijn paradata?

- Achtergrond \Rightarrow kwaliteitsdenken binnen surveyonderzoek
 - » Productkwaliteit
 - » Proceskwaliteit
 - » Kwaliteit organisatie



Wat zijn paradata?

- Couper & Lyberg (2005)
“The ideal situation is that the process variables that we choose are really key so that by measuring and analyzing them we get a chance to check or adjust the process so that the product is delivered according to specifications.”



Wat zijn paradata?

- M. Couper, ASA 1998
 - » Traditioneel survey-onderzoek (PAPI) 'batch-oriented activity'
 - Beperkte mogelijkheden opvolging kwaliteit gedurende het survey proces
 - » Computer assisted survey information collection (CASIC)
 - Groot potentieel aan tijdige real-time data m.b.t. kwaliteit verschillende aspecten van het survey proces

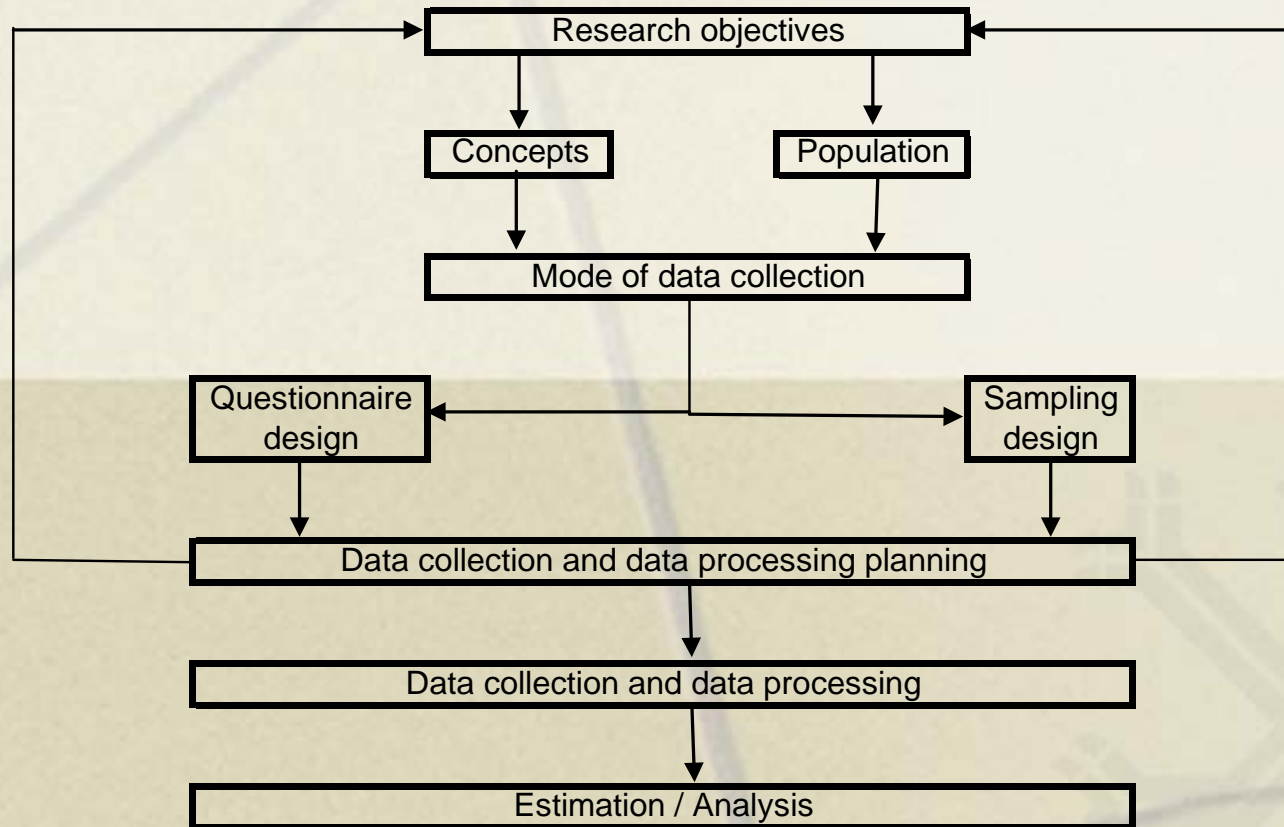


Wat zijn paradata?

- Introductie van computers in het survey proces
 - » Veranderingen in het survey proces
 - » Veranderingen in de evaluatie van het survey proces



Het survey proces



Source: Biemer & Lyberg 2003

Paradata in maten en gewichten

- Fase van ontwerp vragenlijst
 - » Databronnen
 - Keystroke files / trace files / audit trails / websurveys server-side en client-side
 - Eye tracking
 - Audio/video-opname
 - » Analyseniveau procesindicatoren
 - Op niveau steekproefeenheid
 - Op item niveau
 - » Gebruik
 - Navigatie doorheen de vragenlijst o.a. gebruik bepaalde toetsen zoals helpfunctie, back, ...
 - Tijdsmeting – globaal, per vragenblok, per item (response latencies)
 - Interviewergedrag
 - Respondentengedrag
 - Interviewer-respondent interactie



Paradata in maten en gewichten

- Fase van dataverzameling
 - » Databronnen
 - Contactdata vb. per contact nummer interviewer, tijdstip, wijze, resultaat, interactie aan de deur, interviewer perceptie eventuele toekomstige medewerking, buurtkenmerken, indien weigering enkele sleutelvariabelen, ...
 - Inhoudelijke data vb. item non-respons; systematisch antwoordgedrag bij schaalvragen
 - Keystroke files / trace files / audit trails vb. tijdsmeting
 - » Analyseniveau procesindicatoren
 - Globaal op surveyniveau
 - Op steekproefeenheid
 - Per interviewer
 - Op item niveau



Paradata in maten en gewichten

- Fase van dataverzameling
 - » Gebruik
 - Opvolging vooruitgang veldwerk
 - Monitoring interviewer performantie
 - Optimalisering procedures dataverzameling
 - Vergelijking respons en non-responsprofiel op basis van contactdata
 - Planning en evaluatie 'refusal conversion' programma's
 - Responsive designs
 - Meten van nonresponse bias; measurement bias
 - Corrigeren voor nonresponse bias; measurement bias
 - ...



Paradata en uitdagingen voor de toekomst

- Heel wat extra data, welke data wanneer efficiënt en effectief gebruiken?
- Responsive design, evaluatie/bijsturing interviewers \Rightarrow efficiënt en functioneel case management systeem
- Correctie nonresponse bias \Rightarrow samenhang tussen proces variabelen en inhoudelijke variabelen
- Contactdata \Rightarrow geen standaard procedure binnen veldwerkbureaus, kostprijs, kwaliteit, vergelijkbaarheid tussen surveys, cross-nationaal
- Privacy overwegingen
- Toegankelijkheid paradata veronderstelt documentatie en archivering
- ...



Referenties

- Biemer, P. & L. Lyberg (2003). *Introduction to survey quality*. New York: Wiley.
- de Leeuw, E., J. Hox & D. A. Dillman (eds.) (2008). *International Handbook of Survey Methodology*. New York: Lawrence Erlbaum Associates.
- Couper, M. (1998). Measuring survey quality in a CASIC environment. *Proceedings of the survey research methods section, American Statistical Association* (<http://www.amstat.org/Sections/Srms/Proceedings/>).
- Couper, M. & L. Lyberg (2005, April). The use of paradata in survey research. *Proceedings of the 55th Session of the International Statistical Institute, Sydney, Australia*.
- Groves, R. & S. Heeringa (2006). Responsive design for household surveys: tools for actively controlling survey errors and costs. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 169:439-457.
- Heerwegh, D. (2003). Explaining response latencies and changing answers using client-side paradata from a web survey. *Social Science Computer Review*, 21:360-373.
- www.nonresponse.org

