

Survey i en digital tidsalder

Facebook som sampling frame og datakilde

Jonas Toubøl, postdoc,
Sociologisk Institut, Selskab for
Surveyforskning, 24.4.2019

KØBENHAVNS UNIVERSITET



Surveyens tredje æra: Survey + big data

Æra	Sampling	Indsamlingsmetode	Data kontekst
1 (c. 1930-1960)	Tilfældig geografisk afgrænset stikprøve	Personlige dør til dør interviews	Survey står alene
2 (c. 1960-1990)	Tilfældig stikprøve ud fra telefonnummer	Telefon interview	Survey står alene
3 (c. 1990-)	Ikke-tilfældig stikprøve	CAWI	Survey i samspil med big data

Fra Salganik, M. (2017): *Bit by Bit: Social Research in the Digital Age*. Princeton University Press, table 3.1. efter Groves, R.M. (2011) "Three Eras of Survey Research" in, *Public Opinion Quarterly*. Vol. 75(5), 861-871.

Digital udvikling muliggør

- CAWI: Fra interviewer til respondentadministreret indsamling
 - CAWI er PAPI's genkomst
- Rekruttering og sampling ud fra digitale platforme
 - Selv-selekeret deltagelse: ikke-tilfældige stikprøver
 - Annonce baseret rekrutterings genkomst
- Big data og survey er komplementære i mixed methods designs

Repræsentativitet: Ikke-tilfældig sampling

- Tilfældig repræsentativ sampling har problemer
 - Svar raten er for nedadgående = repræsentativitet tvivlsom
 - Vægtning/poststratifikation ofte nødvendig
- Ikke-tilfældig sampling + CAWI = **billig**
- Nok data + post-stratifikation = korrekte estimater

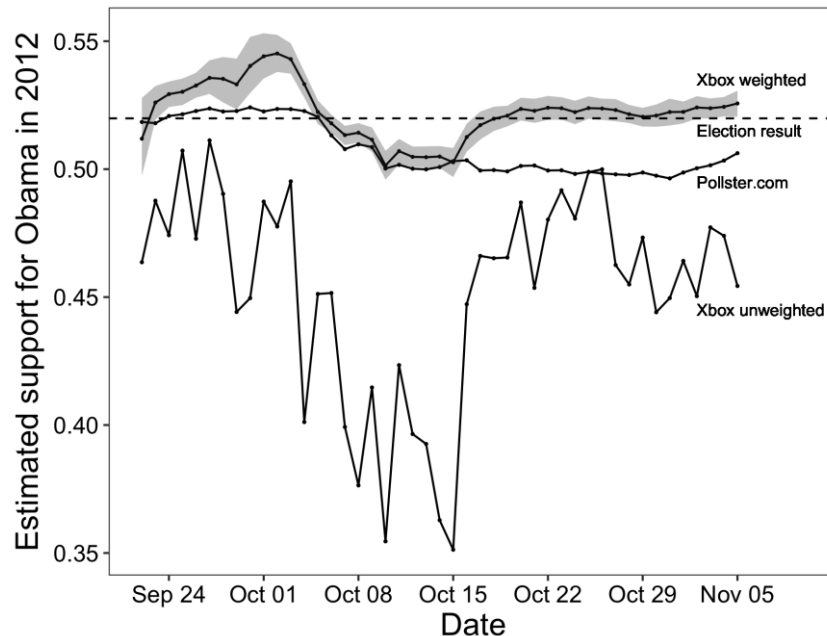


Figure 3.8 fra Salganik, M. (2017): *Bit by Bit: Social Research in the Digital Age*. Princeton University Press på baggrund af Wang W. et al. (2015) "Forecasting elections with non-representative polls" in *International Journal of Forecasting*, vol. 31(2), 980-991.

Måleinstrumentet: Spørgsmål i kontekst

Survey interview classic

Stimulus respons model



Survey interview virkelig

Data samskabes



Måleinstrumentet: Spørgsmål i kontekst

- CAWI giver nye muligheder for at skabe situationer
 - F.eks. billeder, lyd, film, avancerede filtre
- Digital data muliggør at stille spørgsmål i kontekst
 - F.eks. Mobiltelefon data fortæller om aktivitet
 - Stille spørgsmål via sms på bestemte tidspunkter
- Big data muliggør nye populationsafgrænsninger
 - F.eks. registerdata muliggør meget specifikke populationer
- Digitale data giver adgang til observation af kontekst forud for survey
 - F.eks. feltarbejde på sociale medier giver indsigt i sprog og kulturelle koder

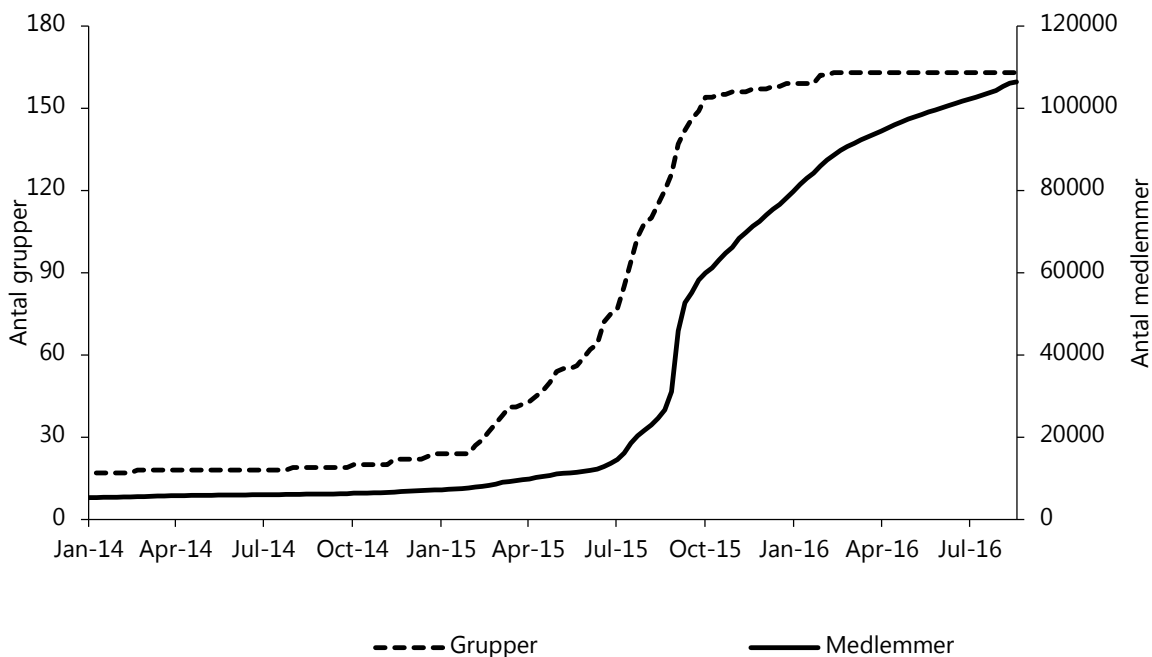
Mixed methods: Survey + big data

Handling {
Mening <- Spurgt med survey
Adfærd <- Observeret med digitale data

- **Beriger hinanden gensidigt**
- Udfordringen er at forene datasæt

Eksempel: Flygtningevenner på Facebook

- Survey af flygtningevenner ifbm. 2015 flygtningekrisen
 - Ph.d. projekt og udført i samarbejde med Peter Gundelach
- Sociale bevægelser er notorisk svære at surveye
 - Svært at forudsige og flygtigt
 - Medlemsskab svært at definere og identificere empirisk



Gundelach, P. & J. Toubøl (forthcoming): "High- and Low-Risk Activism: Differential Participation in a Refugee Solidarity Movement" in, *Mobilization – An International Quarterly*.

Sampling: Facebook grupper og sider

- Population: Flygtningesolidaritetsbevægelsen i Danmark
- Sampling frame: Facebookgrupper og sider relateret til bevægelsen
 - Adgang via administratorer til 287 af 310 (92,6%)
- Rekruttering af stikprøve
 - Opslag i grupper og sider
 - Ens for alle grupper men målrettet på baggrund af forudgående feltarbejde
 - Ofte promoveret af administratorer
 - I flere tilfælde cirkuleret via e-mail lister til også offline medlemmer
 - Selvselekeret medvirken
 - Undersøgelse genstand for diskussion i grupperne
 - Forskeren bør være tilstede

Kvalitative data -> bedre spørgsmål

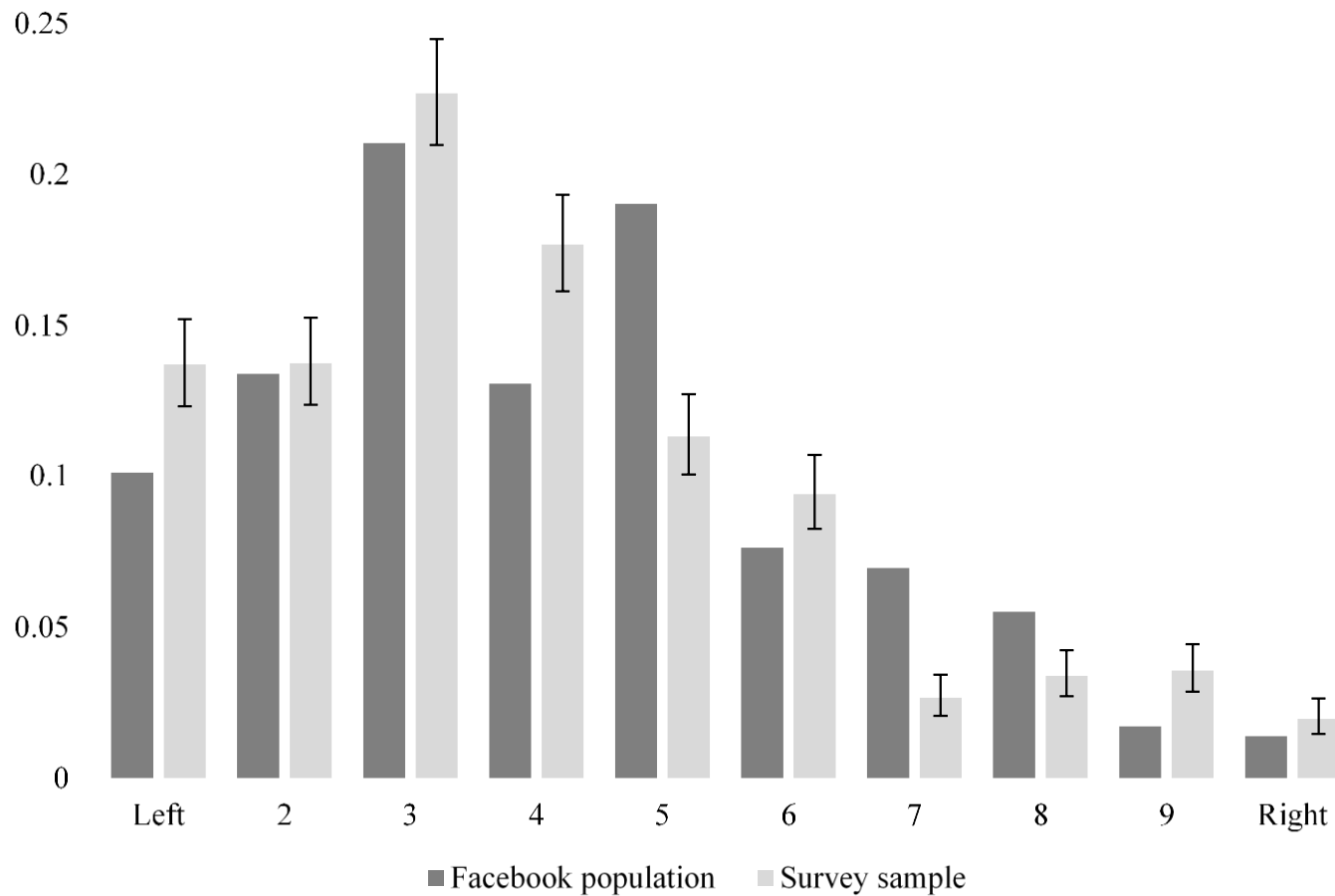
- 51 spørgsmålssider / c. 250 variable
- Resultat: 2.289 valide svar
- Spørgsmål om noget der optager deltagerne
- Spørgsmål med udgangspunkt i kvalitative datakilder og feltarbejde, off- såvel som online

Kombination af data: Digital <-> Survey

- Data fra FB-grupper indsamlet sideløbende
 - Al aktivitet i grupperne
 - Longitudinal individniveau adfærds data
 - Individens adfærd på resten af FB's offentlige sider og profiler
- To links mellem datasæt
 - Gruppe niveau
 - Gruppe specifikke links til online survey
 - Spørgsmål om gruppetilhørsforhold
 - Individniveau
 - Facebook applikation spurgte om tilladelse til at hente og linke profil data
 - (Det kan man ikke mere da Facebook har udnyttet skandaler de seneste år til at lukke adgang til data)

Repræsentativitet

Figure B1. Political orientation of Facebook population and survey sample (95 per cent CI)



Analyse: Gruppekulturens indvirken på individets aktivitet (med Carlsen og Ralund)

- Hypotese: Jo mere konfliktfyldt kommunikation og interaktion i gruppe jo mere politisk protest blandt medlemmerne, alt andet lige
- Forklarende variabel: Fra FB adfærdsdata konstrueres mål for hvor konfliktfyldt gruppernes *stil* er
 - Supervised machine learning
 - 12.000 manuelt kodede posts (6.000 træning/6.000 test)
 - Forudgående online og offline feltarbejde essentielt for korrekt kodning
 - Estimere konfliktniveau i 119 grupper over tid på baggrund af >640.000 poster
- Afhængig variabel: Fra survey data. Individets deltagelse i politisk protest og civil ulydighed vs. humanitær aktivitet

Model

Contentiousness of groups style

Contentiousness of framing

Network {
Personal ties
Organization ties
Civ. Soc. Embed. }

Ind. char. {
Newcomer or veteran
History of activism
Emotions
Biographical availability
Attitudes and values }

Level of participation
in political protest in-
and outside group

```
graph LR; A[Contentiousness of groups style] --> F[Level of participation in political protest in- and outside group]; B[Contentiousness of framing] --> F; C[Network: Personal ties, Organization ties, Civ. Soc. Embed.] --> F; D[Ind. char.: Newcomer or veteran, History of activism, Emotions, Biographical availability, Attitudes and values] --> F;
```

Resultat

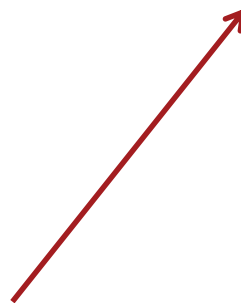
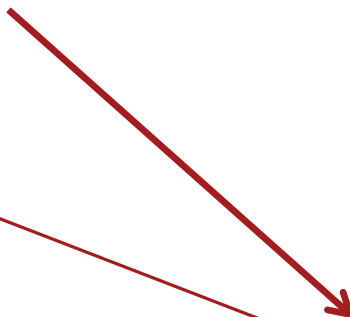
Contentiousness of groups style

Contentiousness of framing

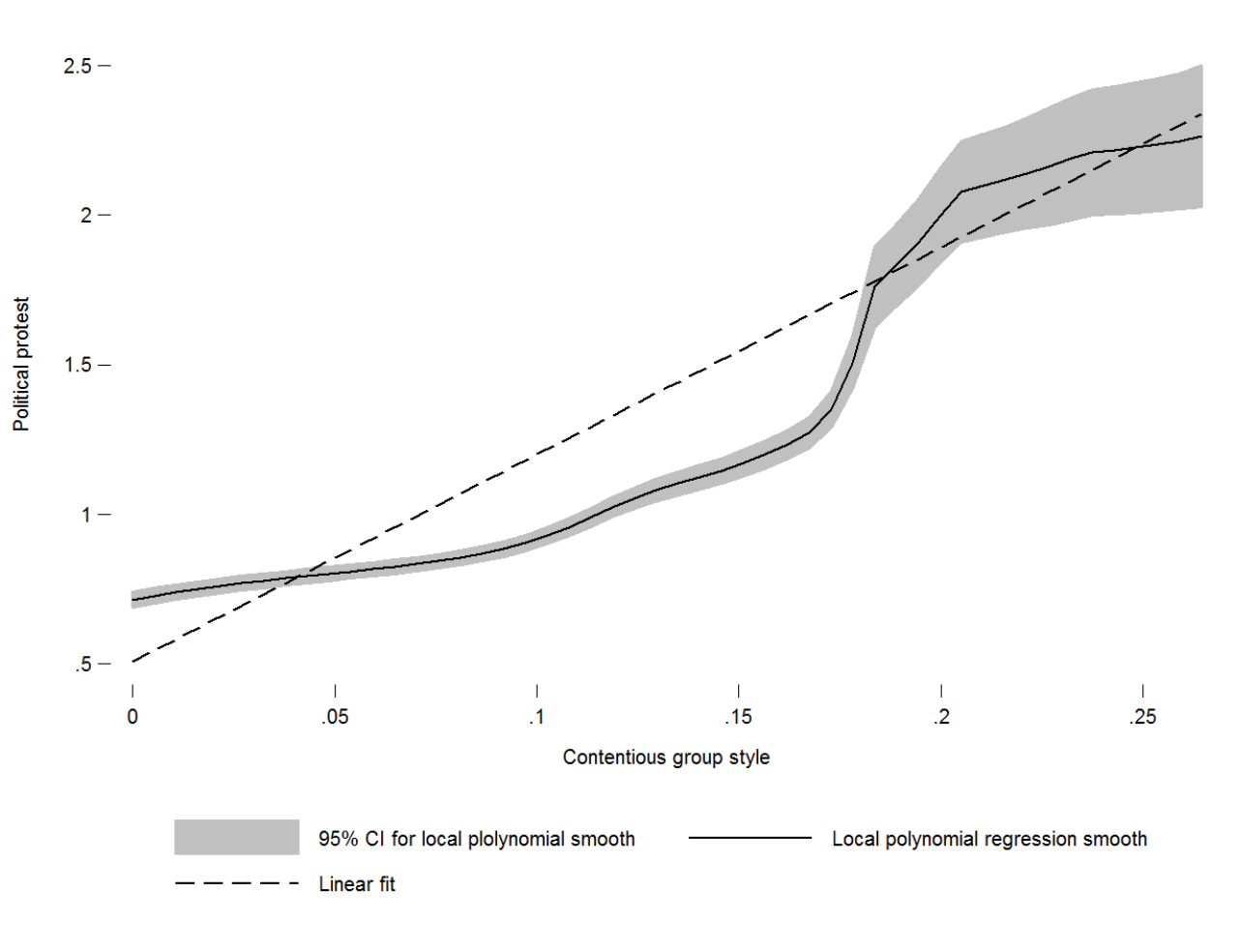
Network {
~~Personal ties~~
~~Organization ties~~
 Civ. Soc. Embed.

Ind. char. {
 Newcomer or veteran
 History of activism
 Emotions
 Biographical availability
 Attitudes and values

Level of participation
 in political protest in-
 and outside group



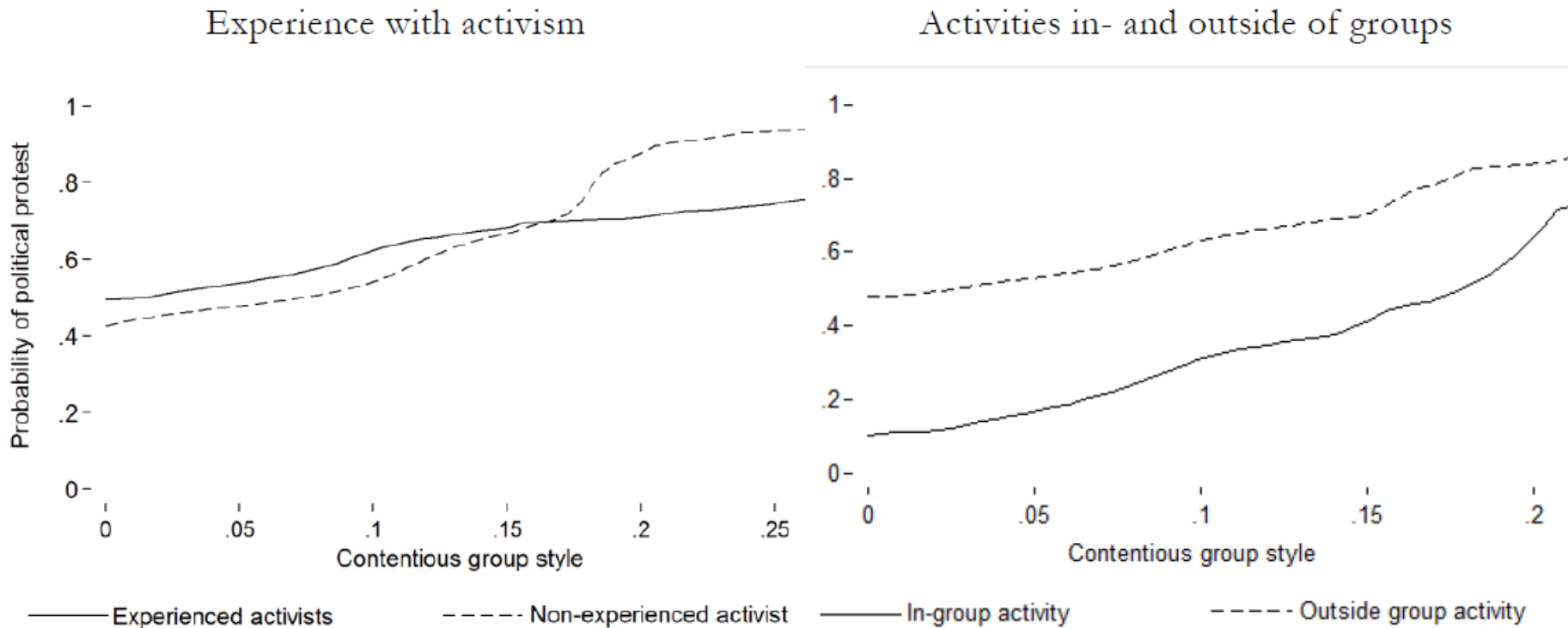
Multilevel model: 75 grupper med >5 observationer



Toubøl, J., H.B. Carlsen & S.E. Ralund (2017): "The consequences of group style for individual participation in political protest: From frame alignment and network to group interaction" in, Toubøl, J. *Differential Recruitment to and Outcomes of Solidarity Activism. Ethics, Values and Group Style in the Danish Refugee Solidarity Movement*. Phd Dissertation. Copenhagen: University of Copenhagen, Department of Sociology.

Aktivisterfaring og aktivitets konteksts medierende effekter

Figure 8.3. Heterogeneity with regard to experience and activities embeddedness in group



Toubøl, J., H.B. Carlsen & S.E. Ralund (2017): "The consequences of group style for individual participation in political protest: From frame alignment and network to group interaction" in, Toubøl, J. *Differential Recruitment to and Outcomes of Solidarity Activism. Ethics, Values and Group Style in the Danish Refugee Solidarity Movement*. Phd Dissertation. Copenhagen: University of Copenhagen, Department of Sociology.

Perspektiver: Sjovere og vildere surveys!

- Big/digital data muliggør nye mindre populationer
 - Ikke-tilfældige samples
 - Big/digital data muliggør bedre poststratifikation
- Digitalisering og mixed methods designs vil forandre survey metode i kvalitativ retning
 - Konteksteffekter bliver aktiv frem for gene
 - Mere eksperimentelle designs
- Vækst i mixed methods designs
 - Survey og big/digital data er komplementære
- Generelle befolknings-surveys vil bestå
 - men indsamles digitalt