Utrecht University

Summer Course Survey Research: Advanced Survey Design

Bella Struminskaya & Peter Lugtig

Department of Methodology & Statistics, Utrecht University

Slides by De Leeuw, Schouten, Hox, Struminskaya, Lugtig

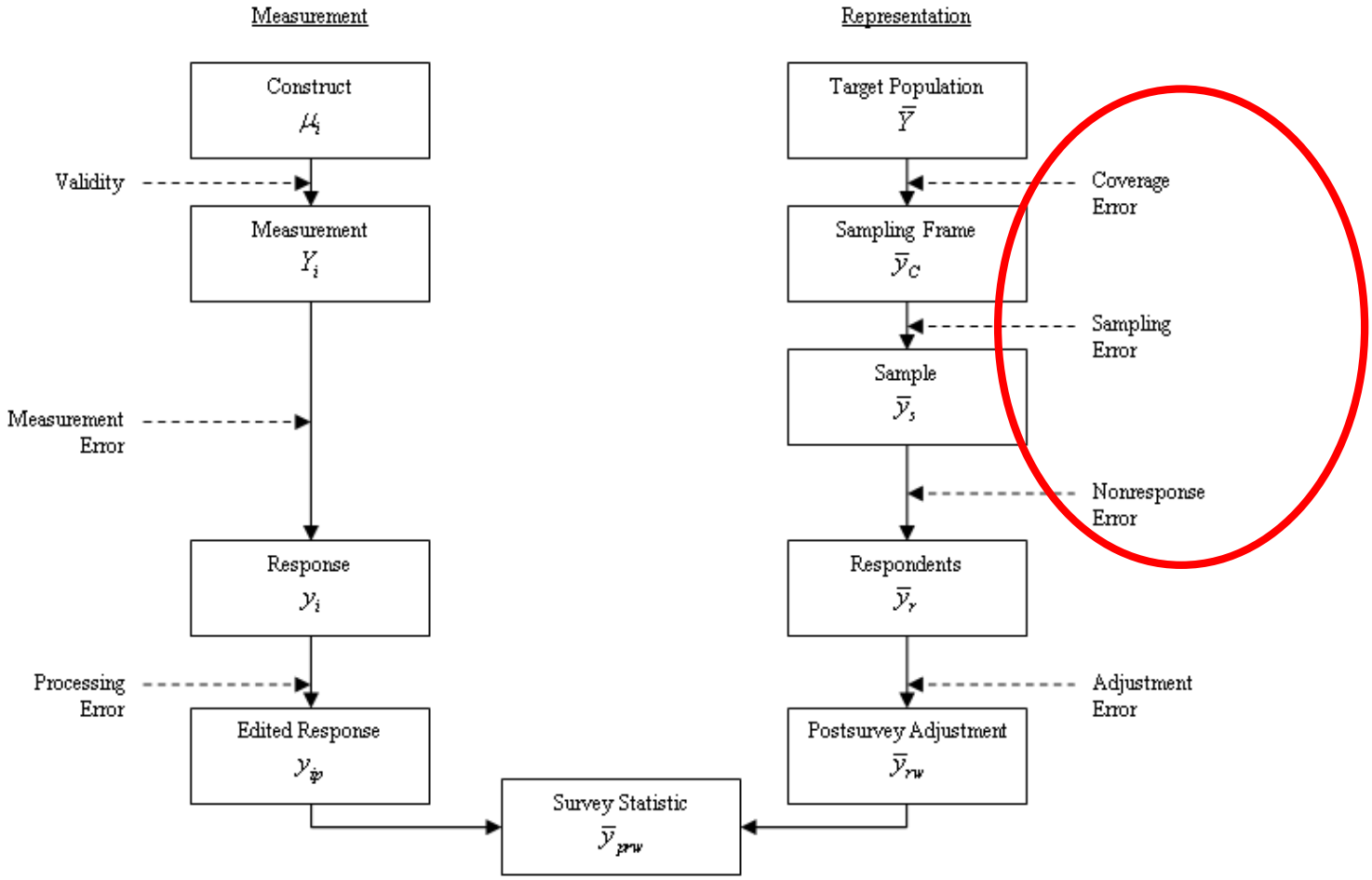
Utrecht University

Sampling frames, recruitment, timing, costs

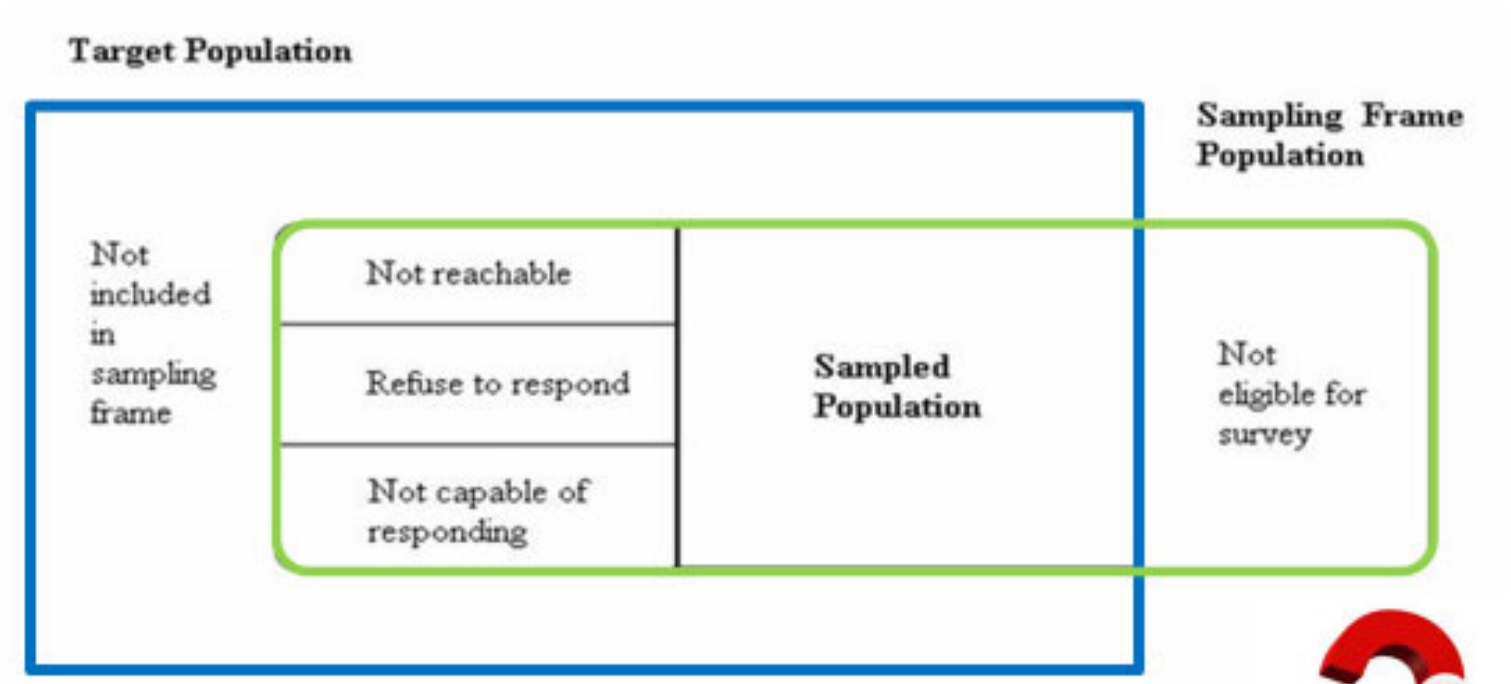
Bella Struminskaya & Peter Lugtig

Department of Methodology & Statistics, Utrecht University

The TSE framework



Coverage error



Can you give an example for each of these sub-groups?

Modes and sampling frames

1. E-mail
   - List only for ‘special’ groups
2. Telephone
   - Landline numbers
   - Mobile phone numbers (no list, random numbers complex)
3. Mail
   - Letters to individuals (population register)
   - Letters to households
4. face-to-face
   - Often first a letter

Modes and sampling frames

1. E-mail
   - List only for ‘special’ groups
2. Telephone
   - Landline numbers
   - Mobile phone numbers (no list, random numbers complex)
3. Mail
   - Letters to individuals (population register)
   - Letters to households
4. face-to-face
   - Often first a letter

Mail sampling frames in Europe

Country	Population register accessible for surveys?	Other sampling frames used by survey programmes
Austria	Yes, Zentrales Melderegister	Address list of the Austrian Postal Services
Belgium	Yes, Rijksregister/ Registre national	Orgassim
Denmark	Yes, CPR	Offentlig Information System (OIS)
Estonia	Yes, Eesti rahvastikuregister	Census, National address register
Finland	Yes, Population Information System	
Germany	Yes (on municipality level), Einwohner-meldeamtregister	Census, GIS areas
Hungary	Yes, Népességnyilvántartó	Census register of dwellings, Area/Settlement
Iceland	Yes, Þjóðskrá	
Italy	Yes (on municipal level), Anagrafe	Household register
Latvia	Yes	Register of Residential Addresses (Office for Citizenship and Migration Affairs), Census, Demographic Statistics Data Processing System
Lithuania	Yes	National address register, Area
Luxemburg	Yes	National Health Insurance database, Administrative and postal address registers.
Malta	Yes	Census-based register, Electoral register, Household register
Netherlands	Yes, Basis Registratie Personen	Postal address register (Cendris Postafgiftenbestand)
Norway	Yes	
Poland	Yes, Pesel	Census areas, OBS statistical sampling frame for social surveys
Slovenia	Yes, Slovenian Central Population Register	
Spain	Yes, Municipal Population Register	Census
Sweden	Yes, Navet/Total Population Register	
Switzerland	Yes	Stichprobenrahmen für Personen- und Haushaltserhebungen
Bulgaria	Not accessible	Census 2011-based register of dwellings
Cyprus	Not accessible	Census-based register, List of addresses from the Electricity Authority of Cyprus (EAC)
Czechia	Not accessible	Census-based register, Czech Statistical Office register of residential addresses, RSO register of Enumeration Districts and Buildings
Greece	Not accessible	Census register of dwellings, Geographical Database
Romania	Not accessible	Census, Multifunctional Sample of Territorial Areas (EMZOT)
Slovakia	Not accessible	Address register, Census register of dwellings, Cadastre register, Telephone register, Geographical database
Croatia	No register	Census-based register, Electoral register, Health register
France	No register	Census, Register of the National Health Insurance, NSEE register of dwellings, Tax register
Ireland	No register	GeoDirectory address registry, Census
Montenegro	No register	Census
North Macedonia	No register	Census
Portugal	No register	Dwelling register, Electricity company (EDP) client frame
Serbia	No register	Census
United Kingdom	No register	Postcode Address File
Albania	Register existing, no information about accessibility	Census
Turkey	Register existing, no information about accessibility	National address register of TurkStat

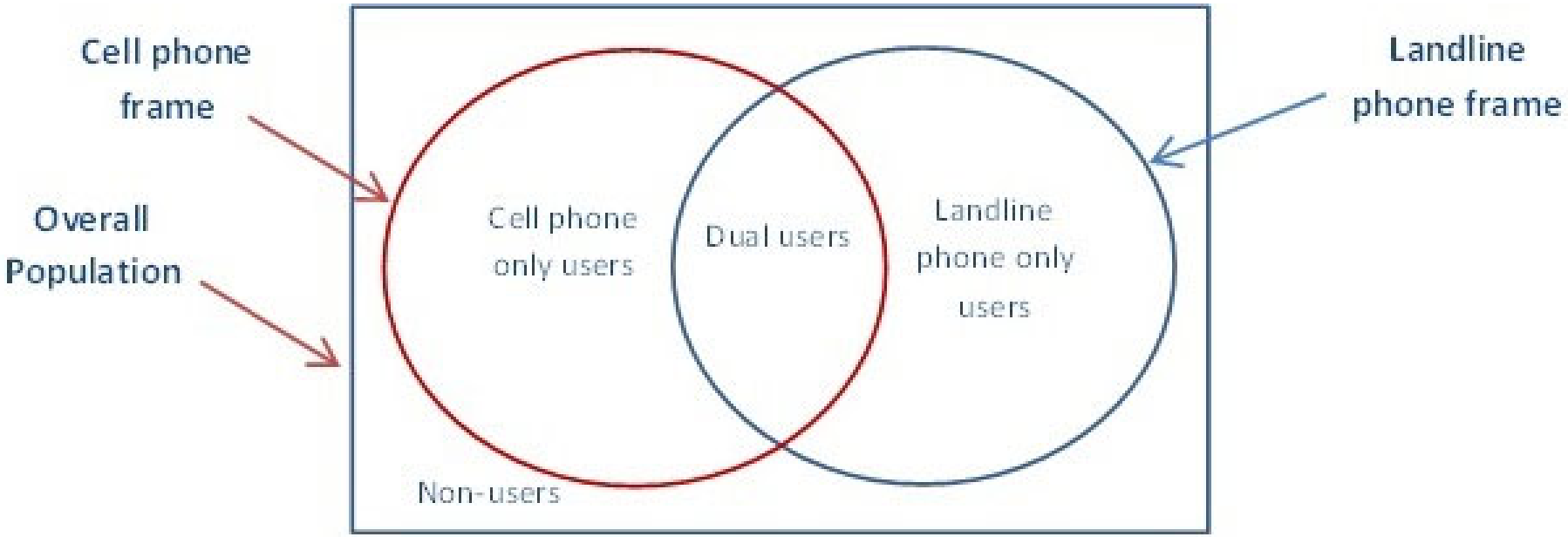
Mail sampling frames in Europe

Country	Population register/ other frame
Austria	Yes, Zentrales Melderegister
Belgium	Yes, Rijksregister/ Registre national
Denmark	Yes, CPR
Finland	Yes, Population Information System
Germany	Yes (on municipality level), Einwohner-meldeamtregister
Hungary	Yes, Népességnyilvántartó
Italy	Yes (on municipal level), Anagrafe
Netherlands	Yes, Basis Registratie Personen
Norway	Yes
Poland	Yes, Pesel
Spain	Yes, Municipal Population Register
Sweden	Yes, Navet/Total Population Register
Switzerland	Yes
Bulgaria	Census register of dwellings, Geographical Database
Greece	Census, Multifunctional Sample of Territorial Areas (EMZOT)
Romania	Address register, Census register of dwellings, Cadastre register, Telephone register, Geographical database
Slovakia	Census-based register, Electoral register, Health register
Croatia	Census, Register of the National Health Insurance, NSEE register of dwellings, Tax register
France	GeoDirectory address registry, Census
Ireland	Dwelling register, Electricity company (EDP) client frame
Portugal	Census
Serbia	Postcode Address File
United Kingdom	No register

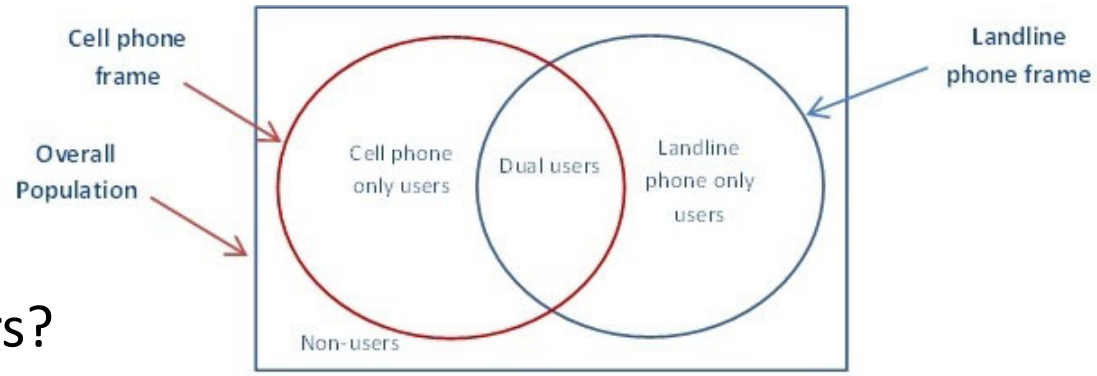
What to do about coverage issues

- Choose the best mode you have
  - Mail as a default
- Multiple contact details for individuals?
  - E.g. mail for most, phone for some
  - Mixed-mode approach
- Multiple lists, overlap uncertain
  - E.g. landline phones and mobile phones
  - Dual-frame sampling
  - See next slide…

Dual-frame sampling



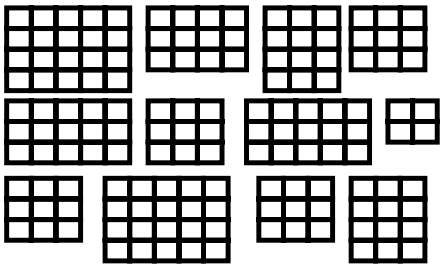
Dual-frame sampling



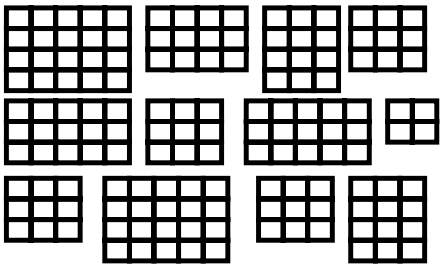
1. Can we estimate the non-users?
   - External data sources
   - Sometimes quality assessment is available
2. Can we estimate the overlap?
   - External sources?
   - Ask respondents:
     - Do you own a cell-phone or landline both, or both?
     - If so, how many of each?

- Note: We can only do this for respondents!
  - People are more likely to respondent if they are easier to contact

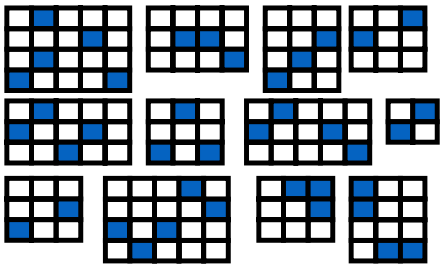
Next step: sampling



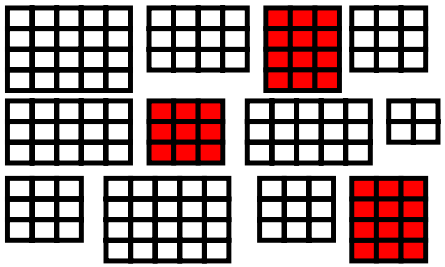
Population of L strata, stratum l contains n_l units



Population of C clusters



Take SRS in *every* stratum



Take SRS of clusters, sample every unit in chosen clusters

Slide based on Balemi (http://www.stat.auckland.ac.nz)

Sampling designs

1. E-mail
   - Special groups-> many covariates on sampling frame
   - simple random sampling or stratification
2. Telephone
   - Little information on sampling frame
   - Simple random sampling
3. Mail
   - Little information address-based, more if from population register
   - stratification
4. face-to-face
   - Quite a bit of information
   - Clustering (costs) and stratification (i.e. multistage)

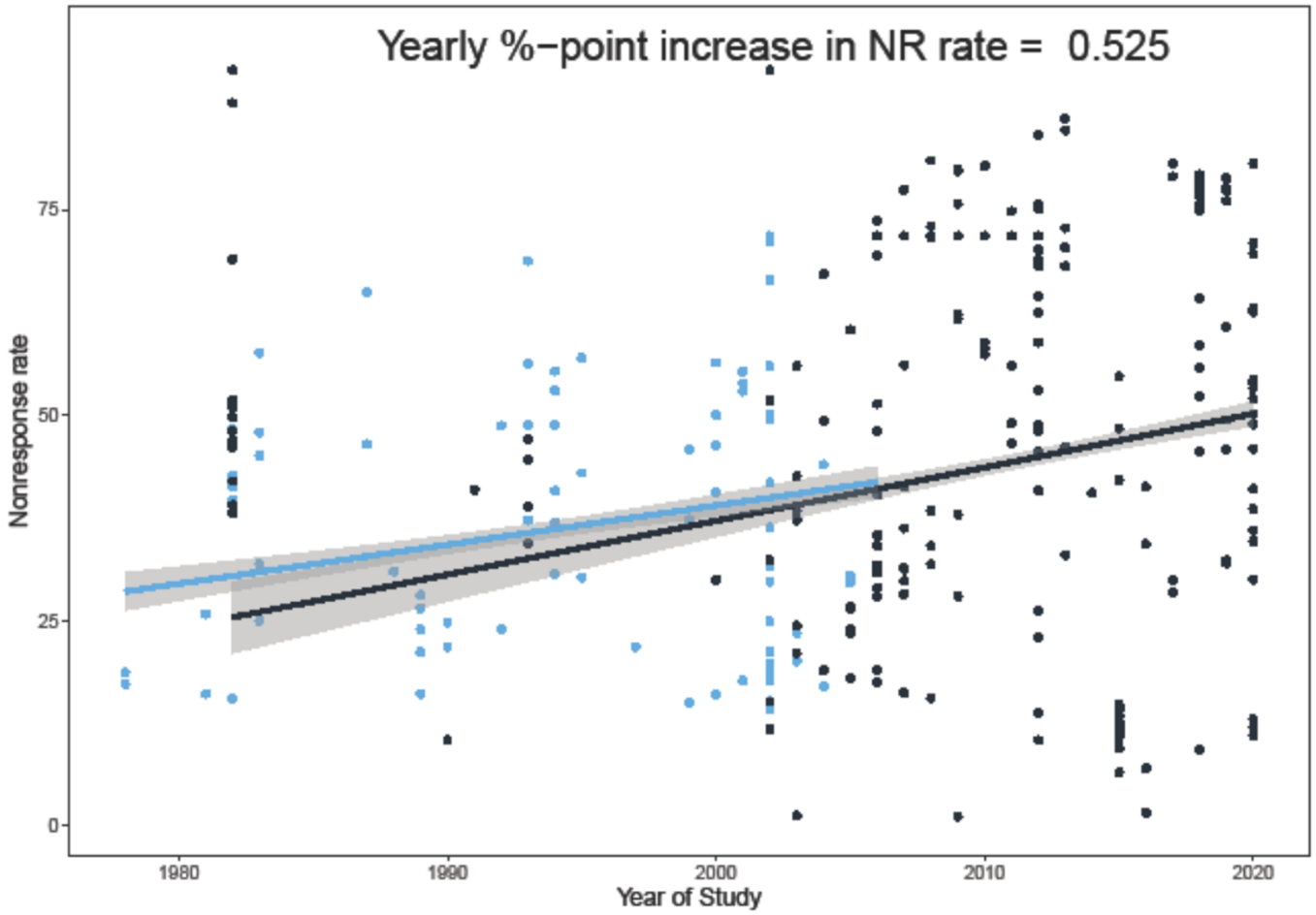
Timing and costs

1. E-mail
   - Cheap
   - Fast (days)
2. Telephone
   - Expensive
   - Quite slow (weeks)
   - Need infrastructure (call center)
3. Mail
   - Moderately expensive
   - Quite slow (weeks)
   - Need infrastructure (mailing + scanning!)
4. face-to-face
   - Prohibitively expensive
   - Very slow (months or years)
   - Need interviewers + IT systems

Nonresponse

- Big issue: response rates are low
- Some groups notorious nonrespondents
  - Men, young or oldest old, lower educated, low political trust
- Most important good practices:
  - Advance letters (well designed, logo etc.)
  - Incentives (unconditional and cash)
  - An interesting topic
  - Multiple completion modes
  - Reminders (ideally in different modes)
  - Refusal conversion

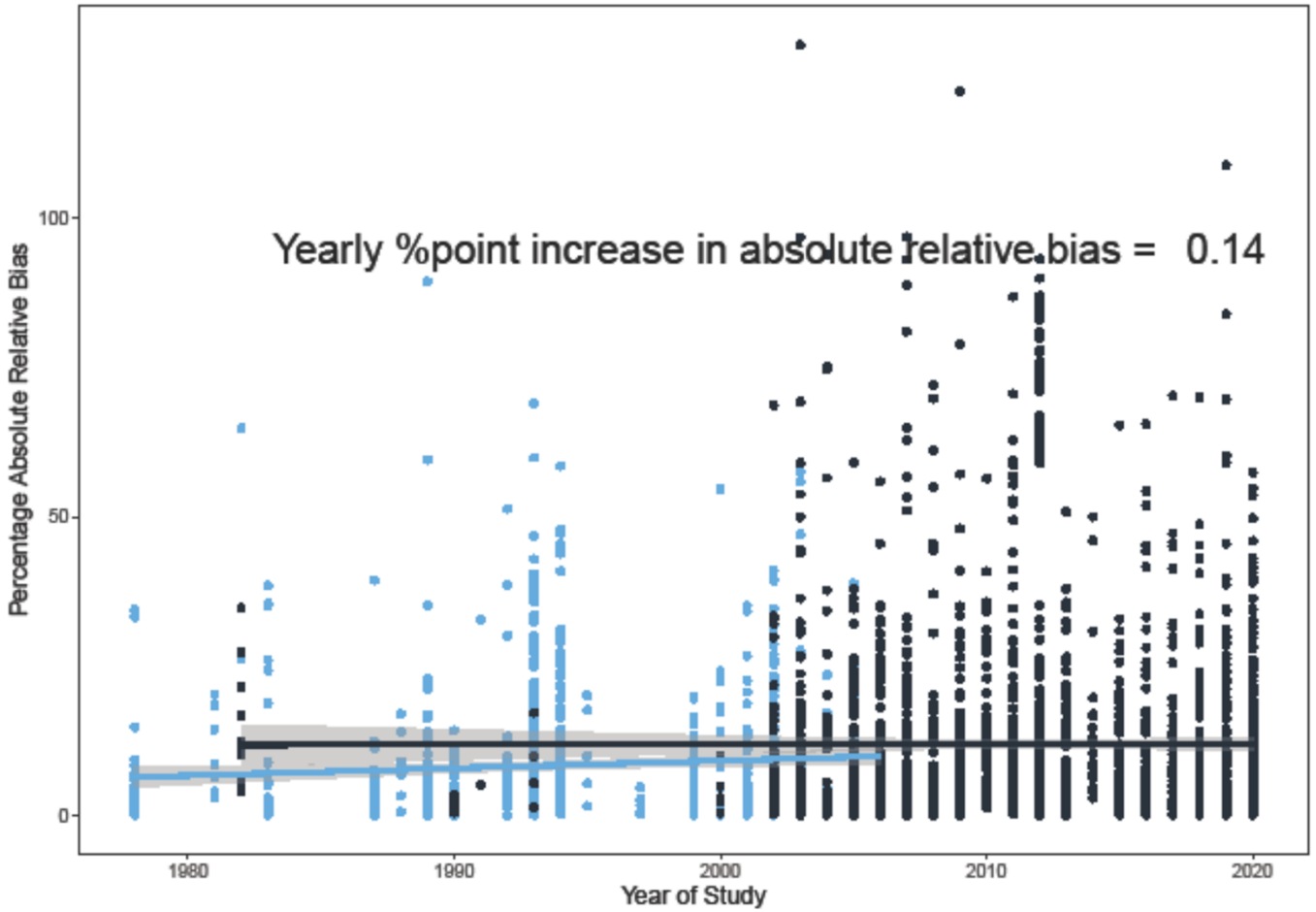
Linear increase in Nonresponse rates



Blue = old articles
Black = new articles

Increase in nonresponse rate is about 5% every 10 years

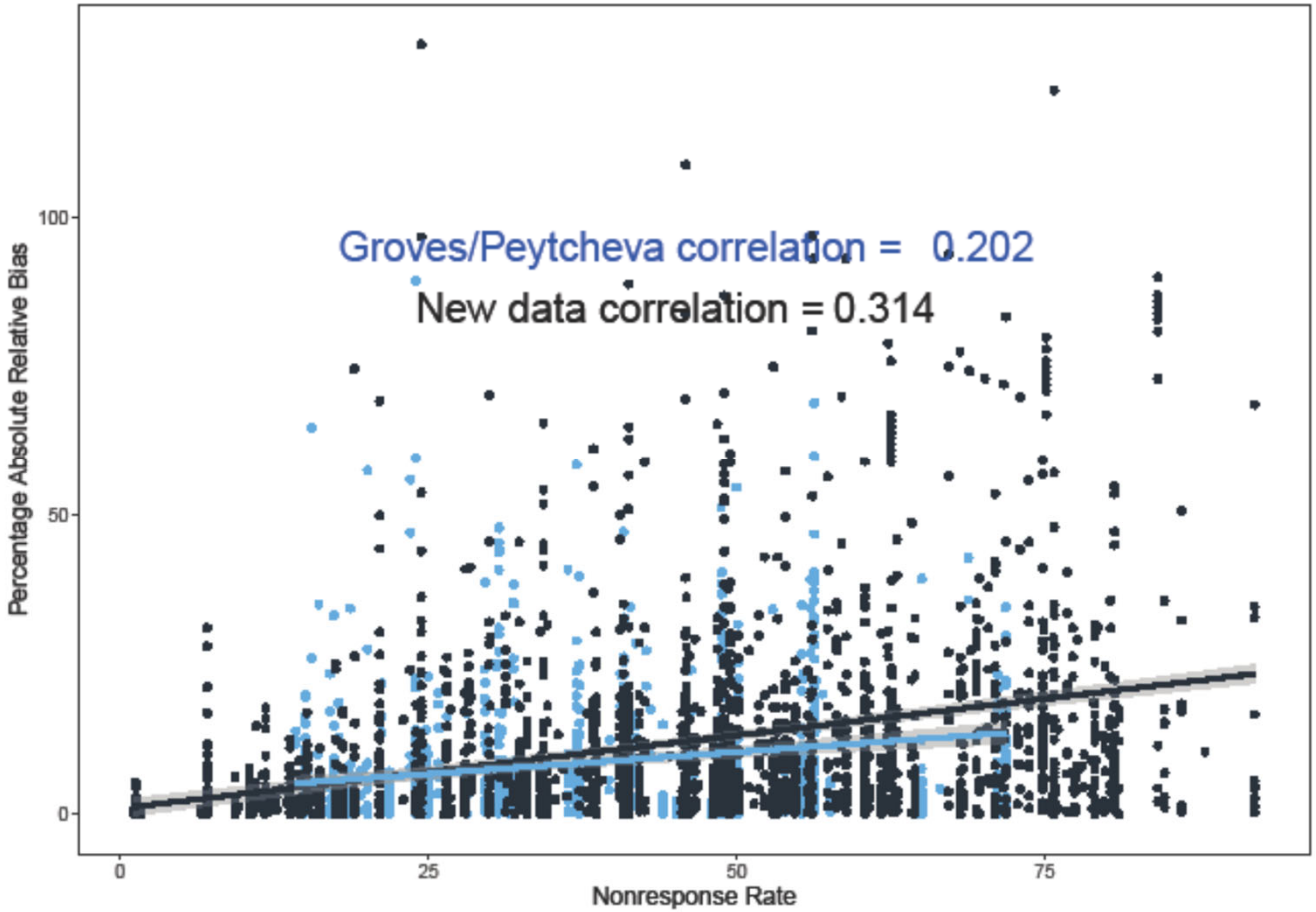
Nonresponse bias almost stable



Blue = old articles
Black = new articles

Increase in bias is about 1.5% every 10 years

Relation between nonresponse rates and bias



Reducing coverage and nonresponse error

1. Think in advance: what groups could I be missing?
   - Coverage and nonresponse
   - What are potential X-variables that relate to R (response)?
2. How does this potentially affect my outcome statistic
   - Missing young people may be ok in study on personality, but bad in study on income
   - What are potential X-variables that relate to Y?
3. Find information about X variables for population
   - E.g. voting behavior, membership of societal organisations, gender, age, level of education, being in work, etc.
4. Build a weighting or imputation model
5. Apply weights (or imputations) in your statistical model

Recommended reading

- Brick, J. M., Dipko, S., Presser, S., Tucker, C., & Yuan, Y. (2006). Nonresponse bias in a dual frame sample of cell and landline numbers. *International Journal of Public Opinion Quarterly*, *70*(5), 780-793.
- Boyle, J. M., Lewis, F., & Tefft, B. (2010). Segmented or overlapping dual frame samples in telephone surveys. *Survey Practice*, *3*(6).
- Demnati, A., Rao, J. N., Hidiroglou, M. A., & Tambay, J. L. (2007). On the allocation and estimation for dual frame survey data. In *Proceedings of the Survey Research Methods Section, American Statistical Association* (pp. 2938-45).