

## **Complicated and expensive: multi-stage sampling of immigrant-origin voters from 11,000 resident registers in Germany**

Presentation handout, Consortium of National Election Studies 25 November 2025, Achim Goerres, University of Duisburg-Essen, co-director of the Immigrant German Election Study, <https://bit.ly/imges2>, [www.achimgoerres.de](http://www.achimgoerres.de)

### **Starting point**

- Assume you want to draw a random sample of German citizens who are 18 or older and who were born abroad (first immigrant generation) or whose parent(s) were born abroad (second immigrant generation); ~ 15 % of all eligible voters OR you want to sample sub-groups thereof
- Germany has no national list of all residents or citizens; but decentralised lists per 11,000 local communities with mandatory residence registration and automatic updating of electoral register
- Organization of local administration decentralized and long-term lack of digital investment: country of birth sometimes in, sometimes not
- Immigrant-origin voters very heterogeneous group with the two most studied groups Russian Germans and those of Turkish descent representing only around 35 %; we know too little about smaller groups, such as Germans of Syrian descent, and in general about third-generation voters who are still minoritised

### **High-quality strategies: unbiased samples of immigrant-origin voters in Germany**

1. Draw random samples from the resident register in high-prevalence contexts, e.g. City of Duisburg with 20 % and more immigrant-origin voters (short interviews with natives, long interviews with immigrant-origin voters)
2. **Mercedes strategy:** Multi-stage sampling of specific groups
  - a. Stage 1: sample of local communities weighted by % of foreigners as a proxy (census data)
  - b. Stage 2: draw random samples of several thousands from the register in each community
  - c. Separate first-name-surname combination from rest of data and use onomastic classification with large databases (Humpert & Schneiderheinze) or newer forms of classification by e.g. n-grammes of word stems (Schnell et al. 2013), then you get a list of “suspects” per local community for each group; add all with dual citizenship and with birth in the respective country
  - d. Interview all or a random sample of your suspects per local community after mail contact
  - e. After a few questions, you know whether you were right; correct classification better for some communities (of Turkish descent) than for others (e.g. Russian Germans) and for those who do not change surnames (more likely among men); name-giving itself a social phenomenon that may be linked to politics
  - f. Challenges & mitigation
    - i. Costs
    - ii. response rates: remedy: door-to-door canvassing, create your own weights by modelling the response rates across contact process
    - iii. onomastic classification quality varies across immigrant and social groups (false-positives are only costly, false-negatives bring bias)

**Alternative strategies:** very biased samples, but useful for much less money, conditional estimates and experiments

3. Sample from number blocs of “ethnic phone contracts” (e.g. “Vodafone 360 Grad Türkei-Flat”)
4. Build database by scrapping names-with-phone-numbers or with e-mail addresses from the internet (e.g. company Data4U in Berlin)
5. Build your own immigrant-origin sample (e.g. Bernd Simon, Uni Kiel)
6. Buy nuanced quota samples from survey providers (e.g. Bilendi, YouGov etc.)