

Register data for policymakers

Jani Erola and Ryan Alberto Gibbons
INVEST Research Flagship, University of Turku

Working with survey data

- Policymakers often rely on survey-based data to develop and evaluate policies
- Research advantages for working with survey data include:
 - Specifically designed for research
 - Allows flexibility to ask questions to directly address specific research questions

Considerable survey drawbacks:

- Non-response
- Memory bias
- Social desirability biases
- Attrition (longitudinal surveys)
- Respondent fatigue

Dropping responses

- Survey drawbacks are compounded by a downward trajectory in survey response rates
- E.g., Britain's labour force survey
 - 1990s - 70% response rate
 - Present - 20% response rate
- Raises further questions
 - What marks out those who choose to respond?
 - What groups are missed entirely?
 - Is it responsible to interfere further in society to collect this information?

What is register data?

- Data on characteristics of (typically) full population of units collected to a storage referred to as *a registry*
- Always includes a unit identifier + information on identified units
- With the help of identifiers, a register can be
 - updated or extended over time
 - matched and merged with other registers on same units using same identifiers

What is register data?

- Units are not always individuals, can be other types of units as well
 - Companies, schools
 - Regions
 - Families
 - etc...

Advantages of working with register data

- Often **not** collected to conduct research.
- However:
 - Data **collection** has **already** been **done**, using society's resources and sometimes interfering in individuals' lives and increasing potential risks to their privacy
 - **Scientific benefits** can come as an additional important reason to downplay potential costs and harms followed from creating a register
 - **No additional interference** from conducting research either - thus supports the freedom of research as well
 - Often **the only reliable source of information**, especially when small subgroups are being studied

Advantages of working with register data

- Accurate information
 - E.g., income from tax registers
 - But also across many demographic, educational, occupational, and health dimensions
- Full population (in the Finnish context, 5 million+)
 - Statistical power
 - Possibility to research multiway associations that are not possible with surveys

Drawbacks of register data

- Data is purely administrative
 - Not designed for research
 - No subjective measures
 - Gaps in the data - e.g., from immigration
- Despite this, there is scope to link other datasets to register data
 - Reduces the need for long questionnaires
 - Allows for longitudinal and life course perspectives beyond survey data
 - Examples:
 - Survey of Health, Ageing and Retirement in Europe (SHARE)
 - Generations and Gender survey
 - INVEST wellbeing survey

Example: daycare attendance in Finland

- Day care attendance less frequent in Finland than in other Nordic countries
- Widely spread assumption among general population: attending daycare before age 4 is harmful for children
- Before 2017: no proper data, no systematic research

So we did it with register data!

Karhula, Aleksi, Jani Erola, and
Elina Kilpi-Jakonen.

”Home sweet home? Long-term
educational outcomes of
childcare arrangements in
Finland”.

*In Childcare, Early Education
and Social Inequality - A Cross-
national Perspective.* Edward
Elgar, 2017.



Home sweet home?

Long-term educational outcomes of
childcare arrangements in Finland

Limits and possibilities of registers

- No direct info on day care attendance in registers
- However, information on home care allowance in family when child age 1-3
- Outcomes:
 - Continuing from comprehensive school to further secondary education by age 17 (non-drop out)
 - General secondary degree at age 20
 - Entry into higher education at age 20

Findings

- Tertiary educated mothers are most likely to choose day care, low-educated home care
- Positive effects for day care, even after controlling for parental education, are weakest for those entering day care at the age of 1
- Effect explained by positive selection: mediated through income and labour market attachment of parents

What followed from the study?

- Extensive public discussion on daycare
- The argument about the harmfulness of daycare attendance mostly dropped
- Policy discussions moved from effects on children to benefits for families