USING ONLINE GENEALOGICAL DATA FOR DEMOGRAPHIC RESEARCH: AN EMPIRICAL EXAMINATION OF THE FAMILINX DATABASE

Andrea Colasurdo[1], Riccardo Omenti[2]

[1]Kinship Inequalities Research Group, Max Planck Institute for Demographic Research

[2]Faculty of Statistical Sciences, University of Bologna

MAX PLANCK INSTITUTE FOR DEMOGRAPHIC RESEARCH

BACKGROUND

Online genealogies represent a novel data sources for demographic research.

❑ **Promises**

- Analyze demographic processes in periods and countries without accurate data
- Transnational kinship networks

❑ **Pitfalls**

- Numerous biases and selectivity issues
- Many missing values in main demographic variables

➢ Employed in several studies without properly addressing its selectivity and data quality issues .

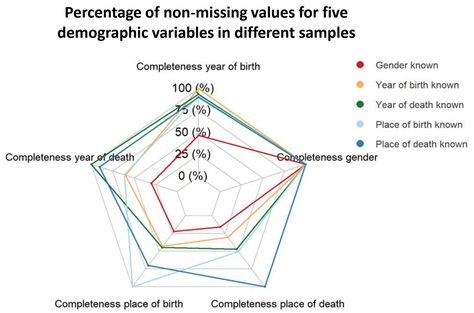
OBJECTIVES

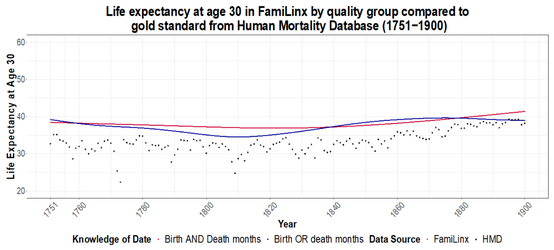
- Evaluate the **quality** and the **completeness** of demographic variables at both the individual and the family network level in a big genealogical digital database.
- Analyze the impact of completeness and quality of demographic information on age-sex distribution and demographic estimates from online genealogical populations.

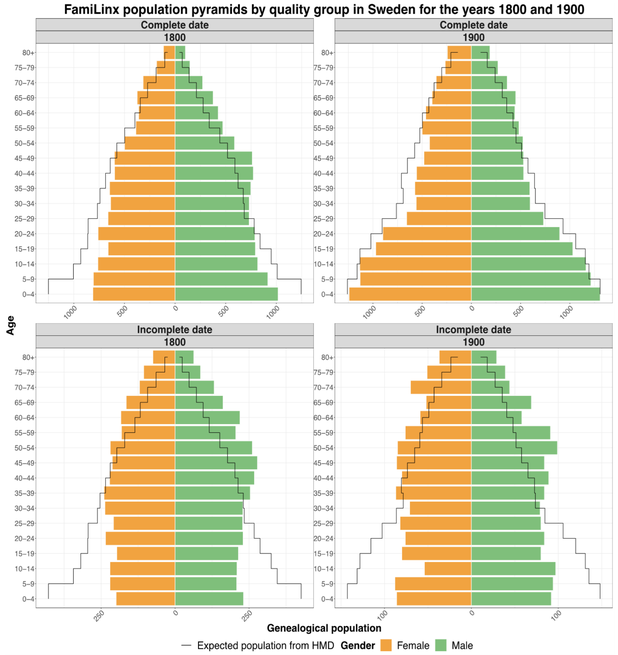
DATA: *FamiLinx*

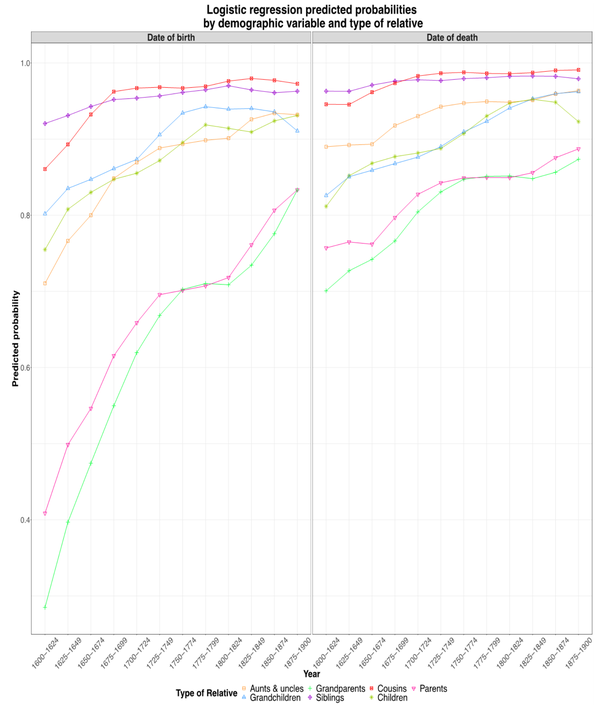
- Database with 86 million unique profiles from the last 400 years.
- Vital events mainly concentrated in the Global North.
- Subsample of approx. 7 million "Focal" individuals with at least **one recorded parent or child** and with **birth and death events** within the 20 most represented countries.

RESULTS









DISCUSSION

- Individuals from FamiLinx with demographic information of higher quality and completeness represent a highly selected group with higher survival and (likely) higher SES.
- Representativeness of FamiLinx populations improves towards the end of 19th century.
- Completeness and quality of demographic variables are clustered within kinship networks.
- FamiLinx offers much promises to study the transmission of demographic behaviors among focal individuals and their relatives.
- The implementation of bias-adjustment methods and a careful sample selection are needed.