ICS PhD Project 1:
Understanding fertility outcomes by predicting them

Aim
To get a better understanding of fertility outcomes (i.e. how many children people want and have) using theory-driven methods, machine learning, and a data challenge.

Theoretical background
Sociological research has produced a sophisticated body of work on the many characteristics associated with fertility outcomes—how many children people have and when. These range from the social environment during upbringing to partnership trajectories in later life, from social interactions with friends to family policies in society, and from biological differences to differences in values. Despite these advancements, there is no systematic way in which the relative importance of these characteristics can be quantified, hindering cumulative science. Moreover, these characteristics typically only explain a minor fraction of variation in fertility outcomes. This is exemplified by the failure of current models to predict even short-term fluctuations in fertility outcomes.
When models are not predictive of the real world, this could be because of flawed theory, but also because of poor data and measurement, suboptimal analytical strategies, inherent unpredictability, or a combination of these. Distinguishing between these causes is important for the study of any trait. This project uses a combination of theory-driven and data-driven methods to be able to highlight the most important causes of failure to predict fertility outcomes, and thereby advance our understanding of them.
The PhD project will focus on establishing predictive benchmarks for fertility outcomes, based on existing theories of fertility and on machine learning techniques harnessing the power of many variables. In addition, a dedicated data challenge is organised in which teams compete to make the best predictions of fertility outcomes. Differences between theory- and data-driven methods in their predictive ability gives unique understanding on the causes of poor prediction. This PhD project is part of the "Understanding fertility outcomes by quantifying the (un)predictable" project (VIDI grant awarded to Gert Stulp).

Research design
This project will draw on existing survey data that contain large numbers of variables (i.e., the LISS panel and the GGS). For both datasets, predictive benchmarks will be established based on a) existing theories of fertility and conventional statistical methods, and b) data-driven machine learning methods. Through an international data challenge (or scientific mass collaboration), upper limits to predictions are set for the LISS panel survey data. This is an interdisciplinary project encompassing sociology, demography, and data science.

Profile
- A degree in quantitative social science or in data science with affinity to the social sciences
- Experience with programming language R and/or python
- Experience with machine learning techniques and statistical methods for large dataset or related experience and a drive to learn such methods
- Aptitude to plan and coordinate a major project (i.e. the data challenge)

Literature
- Balbo, N., Billari, F. C., & Mills, M. (2013). Fertility in advanced societies: A review of research. European Journal of Population, 29(1), 1–38. https://doi.org/10.1007/s10680-012-9277-y
- Salganik, MJ, et al. (2020). Measuring the predictability of life outcomes with a scientific mass collaboration. Proceedings of the National Academy of Sciences, 117(15), 8398–8403. https://doi.org/10.1073/pnas.1915006117

Project initiators
Gert Stulp (project leader, RUG), Malvina Nissim (RUG), René Veenstra (RUG)

Location
Department of Sociology, University of Groningen