

International Journal of Population Data Science

Journal Website: www.ijpds.org



Are births predictable with linked survey and register data? Evidence from the Predicting Fertility data challenge

Elizaveta Sivak¹ and Gert Stulp¹

¹University of Groningen, Groningen, Netherlands

Accurate predictions of life outcomes can inform social theory and policy. Yet, social science predictions often perform poorly. One common explanation is small sample sizes of social datasets. We examine the predictability of having a child within three years under conditions that are close to the best currently possible. We use full-population Dutch registers and LISS survey data, linkable to the registers, in the Predicting Fertility data challenge. LISS data includes a wider range of theoretically relevant variables, including subjective measures such as fertility intentions. Register data provides vast, high-resolution coverage of life-course trajectories but lacks subjective measures. This unique framework enables us to leverage the strengths of these datasets to assess the current predictability. Over 150 people participated in the data challenge

and submitted over 70 models, both traditional machine learning and cutting-edge approaches. Predictive performance remained modest for both datasets: even the best models fell below the upper limit of predictability caused by randomness in conception and fetal survival. Survey-based predictions performed slightly better. Combining the best survey- and register-based models – by using the predicted probabilities from the best register-based model as an additional feature in the best survey-based model – slightly improved predictions. These results suggest that accurately predicting individual fertility remains highly uncertain, even in the short term and with extensive data. Modest predictive accuracy is unlikely to stem primarily from limited sample size, but may reflect the inherent unpredictability of life outcomes.

