



Fertility prediction challenge, Episode I: does survey data beat population registries?

ODISSEI Conference for Social Science
in the Netherlands 2023

Elizaveta Sivak, Gert Stulp, Javier Garcia-Bernardo,
Malvina Nissim, Tom Emery, Adriënne Mendrik,
Paulina Pankowska



university of
 groningen



ODISSEI



Eyra



CENTERDATA
research institute



- Research on fertility behaviour: many factors identified; still, our knowledge might be incomplete

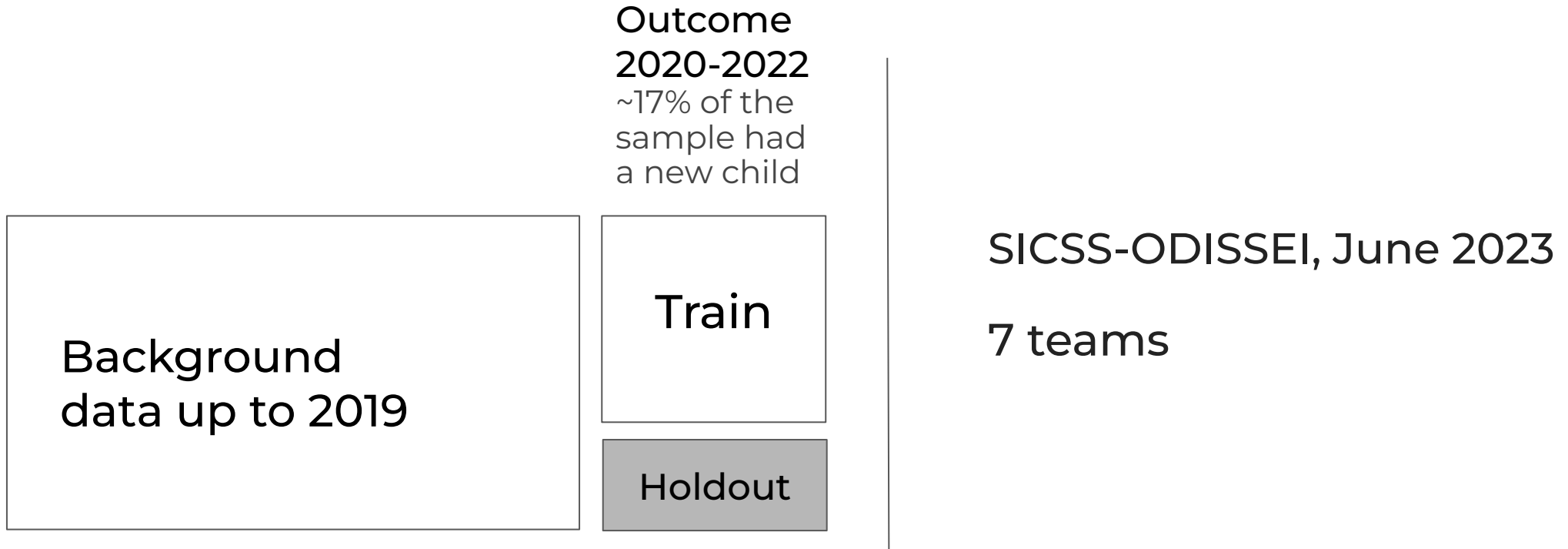
A data challenge to predict fertility outcomes

- assess current upper levels of predictability
- highlight the gaps in our current knowledge and get insights



A pilot

The task: to predict who will have a child in the next three years (2020-2022)



SICSS-ODISSEI 2023 participants: THANK YOU



Dianna Amasino



Vardan Barsegyan



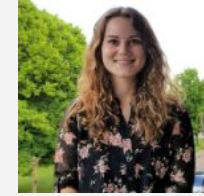
Emily Cantrell



Santiago
Gómez-Echeverry



Hofland Maud



Laura Jansen



Sanni Kuikka



Marie Labussière



Francesco Marolla



Huyen Nguyen



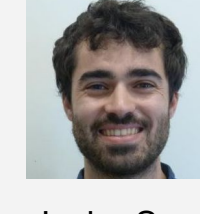
Alon Pertzikovitz



Samuel Plach



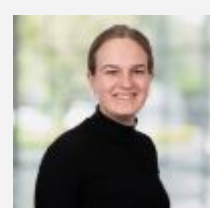
Žiga Puklavec



Javier San Millán
Tejedor



Hanna Schleihauf



Lotte Schrijver



Lucas
Spierenburg



Ely Strömberg



Tomas
Turner-Zwinkels



Tony Wei-Tse
Hung



Xiao Xu



A pilot

Phase I:

LISS panel: all Core studies +
Background (from 2007-2019)



Submission platform to increase
computational reproducibility



Phase II:

Dutch population registries (CBS):
a preprocessed dataset with mostly
2019 data + networks + external data
about municipalities



Evaluation metric: F1 score

Wide data (survey) vs long data (registries)?

Your bets:

Survey data

vs

Registry data

Thousands of variables

~100 variables

Your bets:

Survey data

vs

Registry data

Thousands of variables

~100 variables

N = 1200

N = 6 000 000

Your bets:

Survey data

vs

Registry data

Thousands of variables

~100 variables

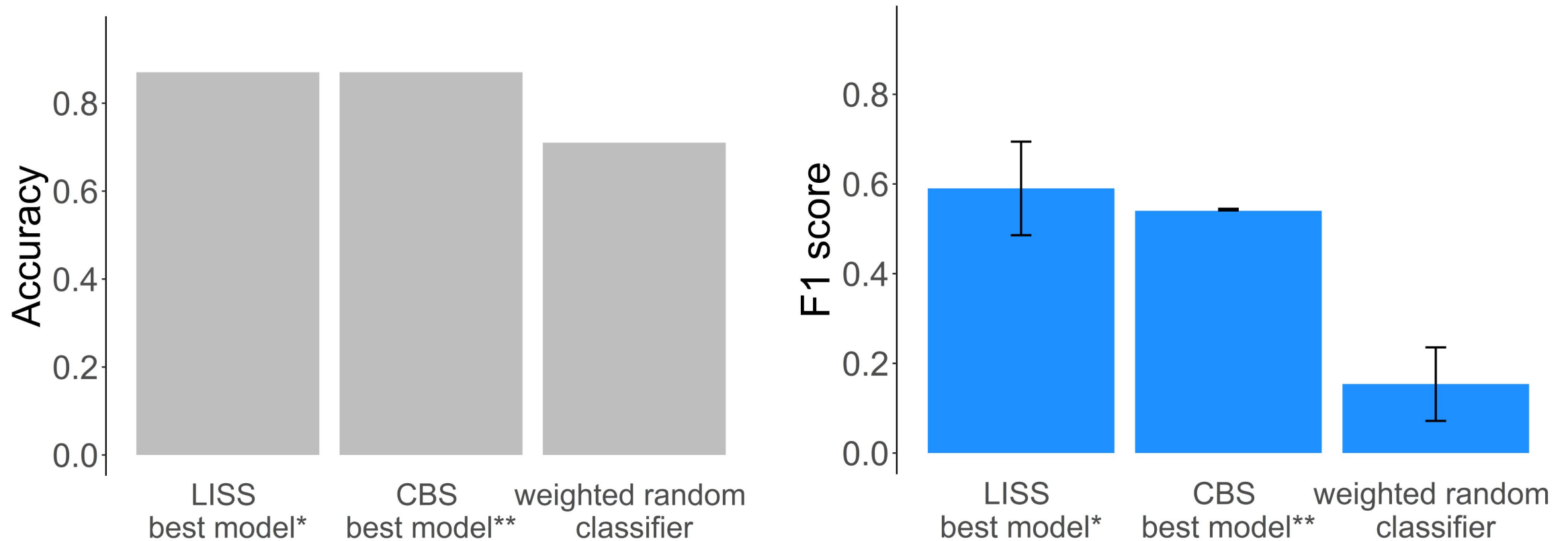
N = 1200

N = 6 000 000

Values, preferences,
attitudes, beliefs

No such variables

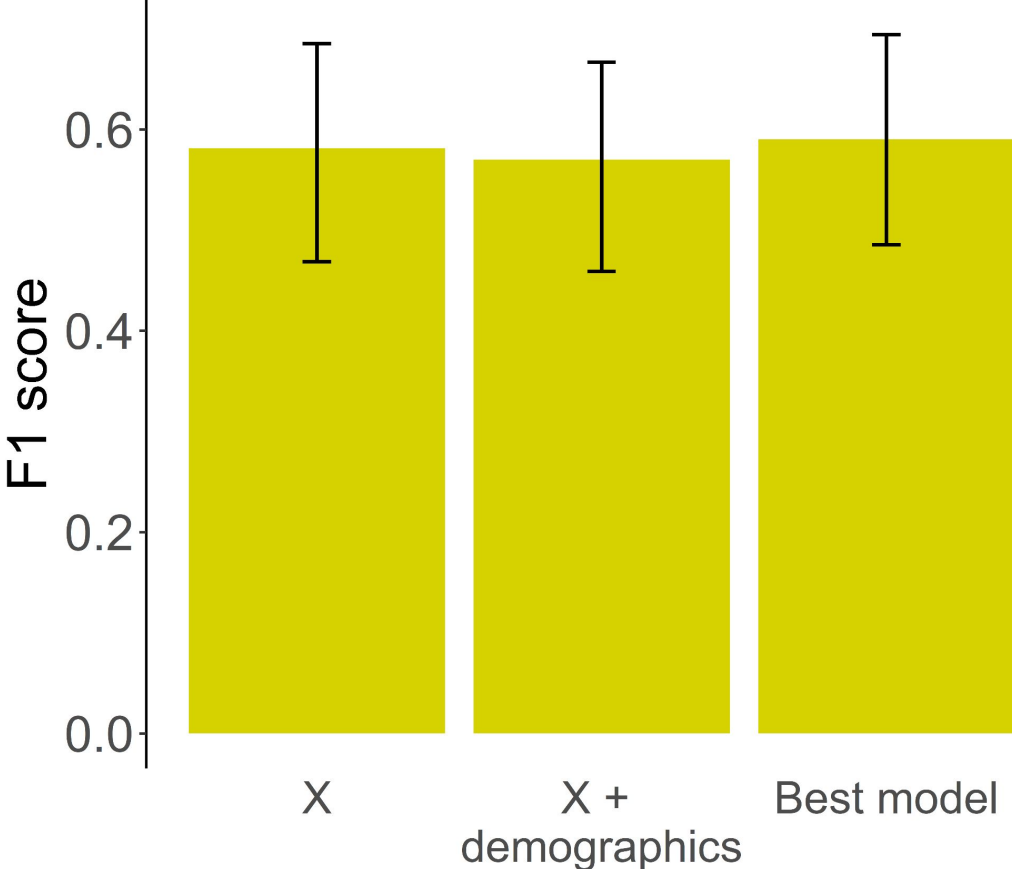
Predicting who will have a child in the next three years



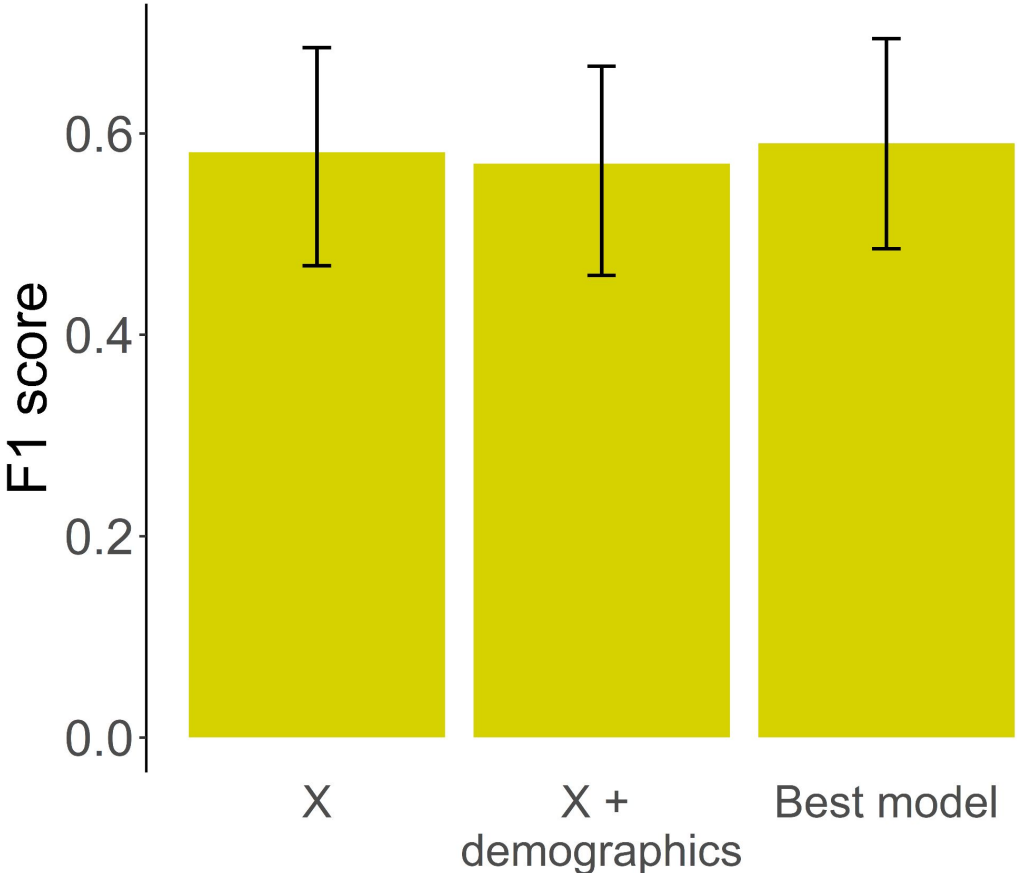
* League of Lager team

** STL Trio Titans team

LISS: what worked



LISS: what worked



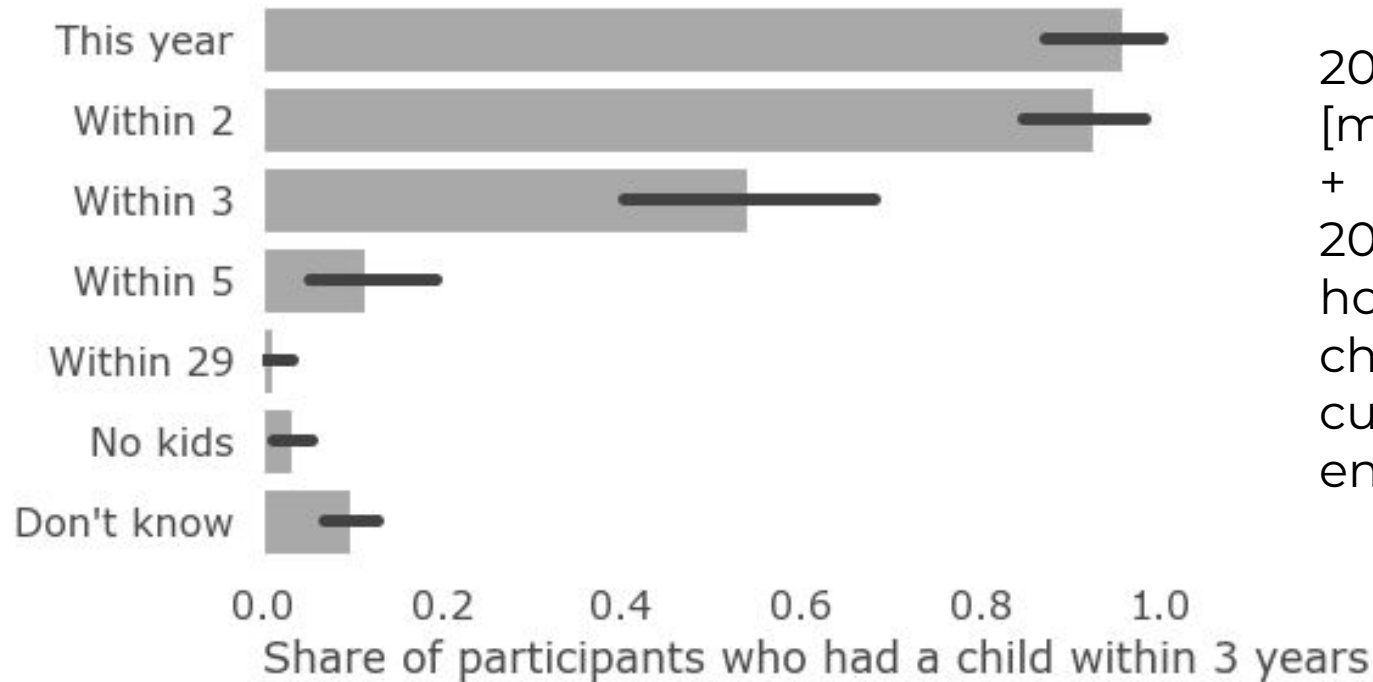
X = Fertility Intentions

2019: Do you think you will have [more] children in the future?

+

2019: Within how many years do you hope to have your [first / next] child? If you [your partner] are currently expecting a child, please enter a zero (0)

Fertility intentions and actually having a child within 3 years

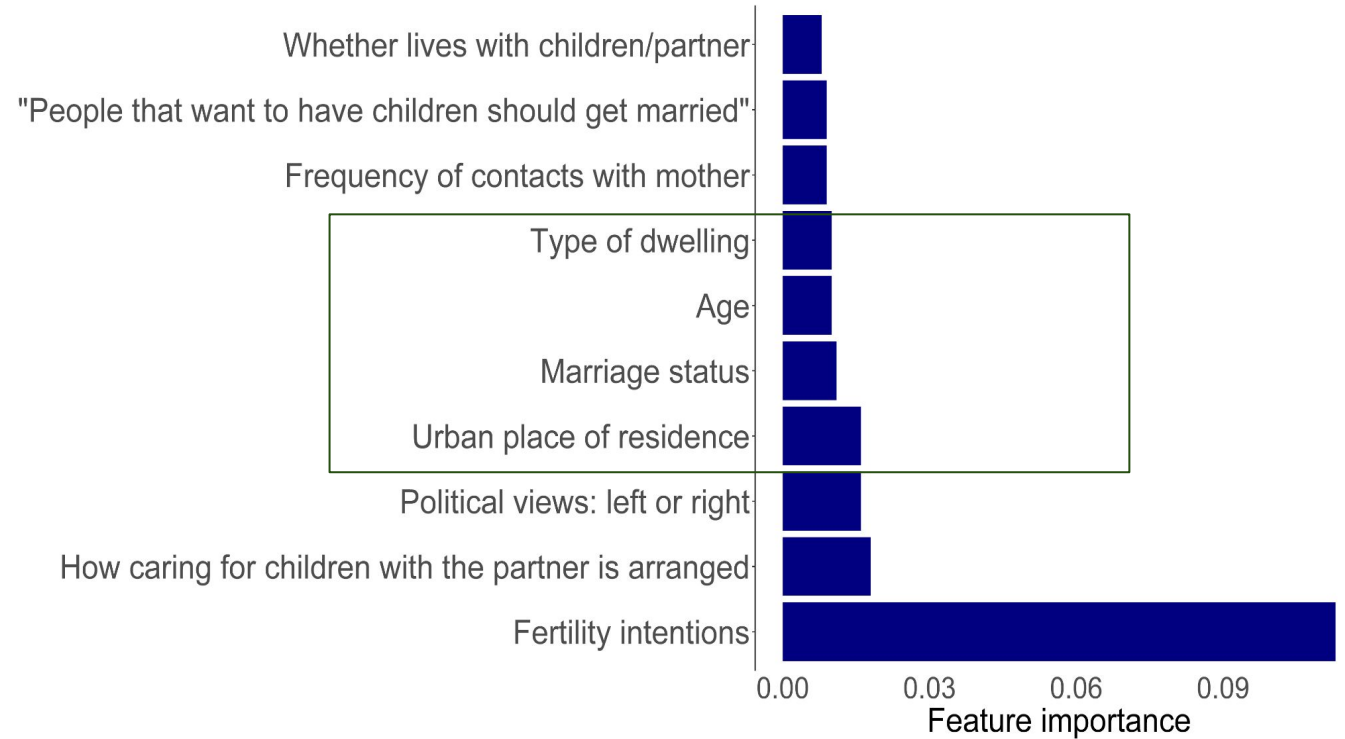
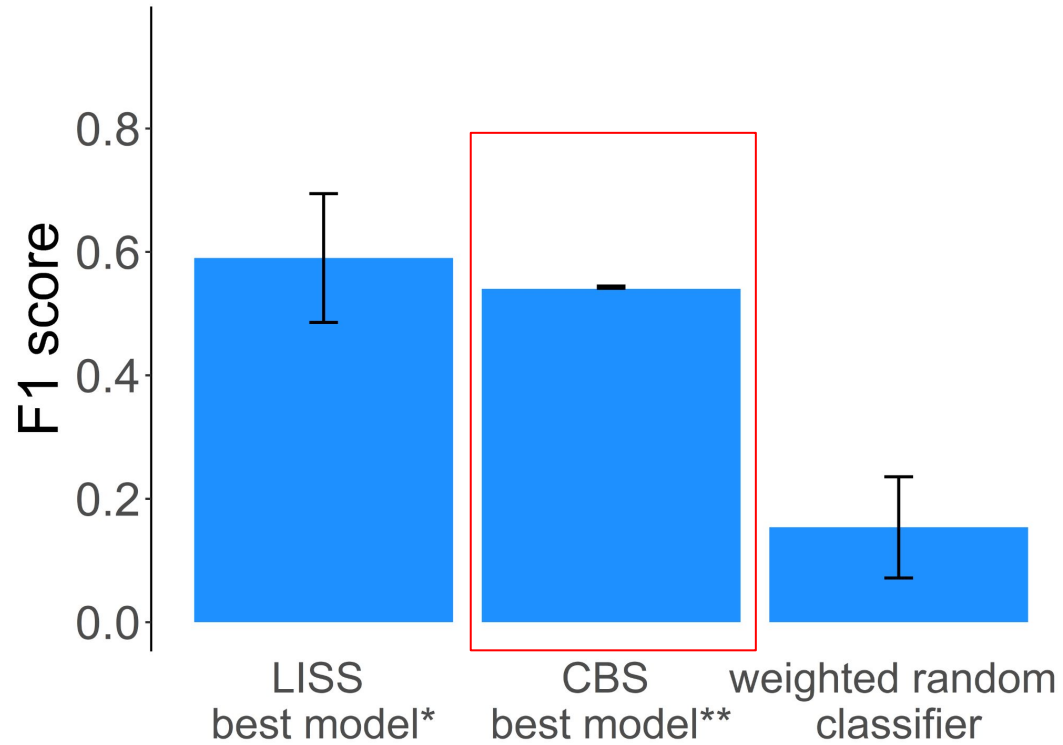


2019: Do you think you will have [more] children in the future?

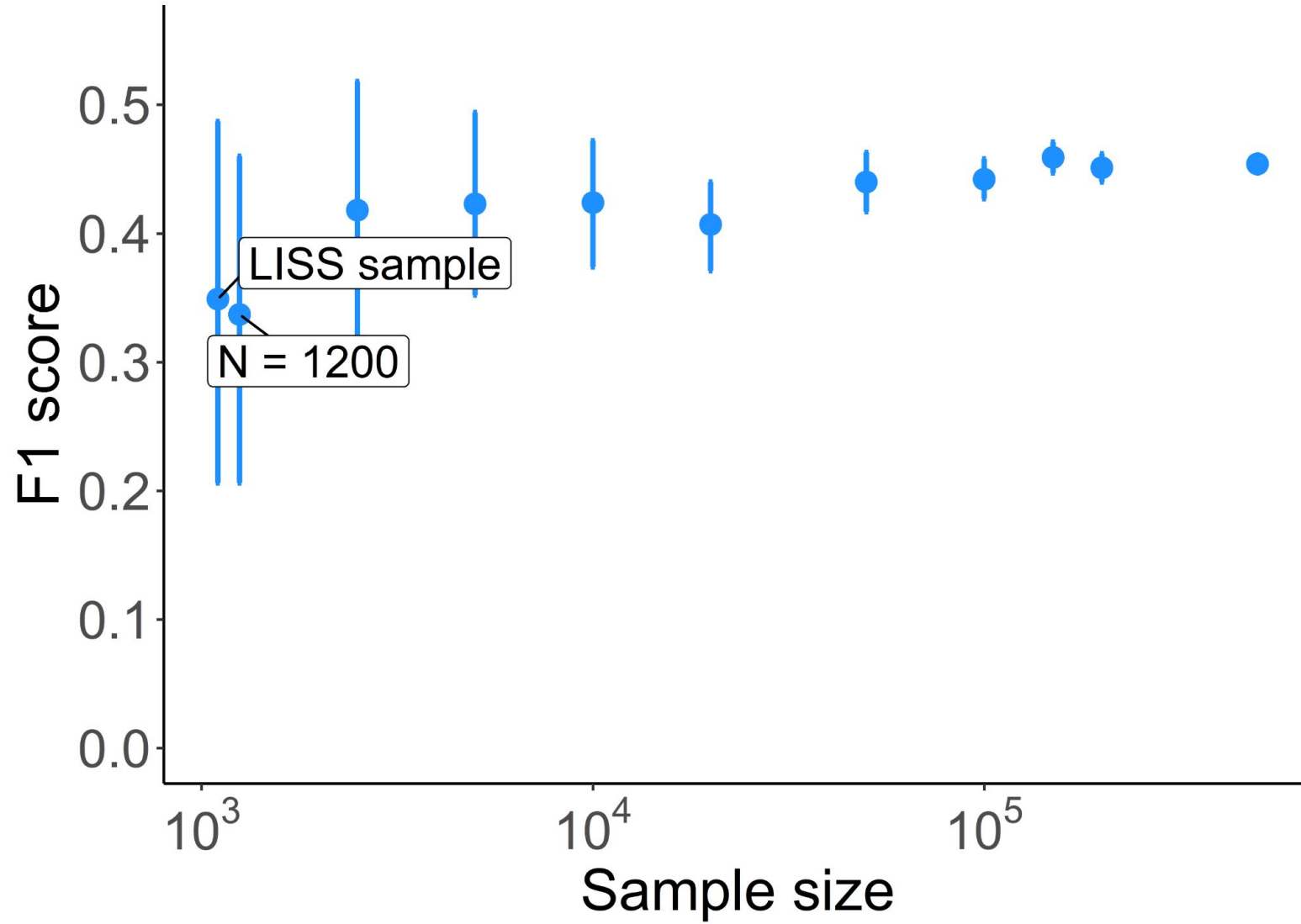
+

2019: Within how many years do you hope to have your [first / next] child? If you [your partner] are currently expecting a child, please enter a zero (0)

CBS data: what worked?



CBS data: what worked





pilot

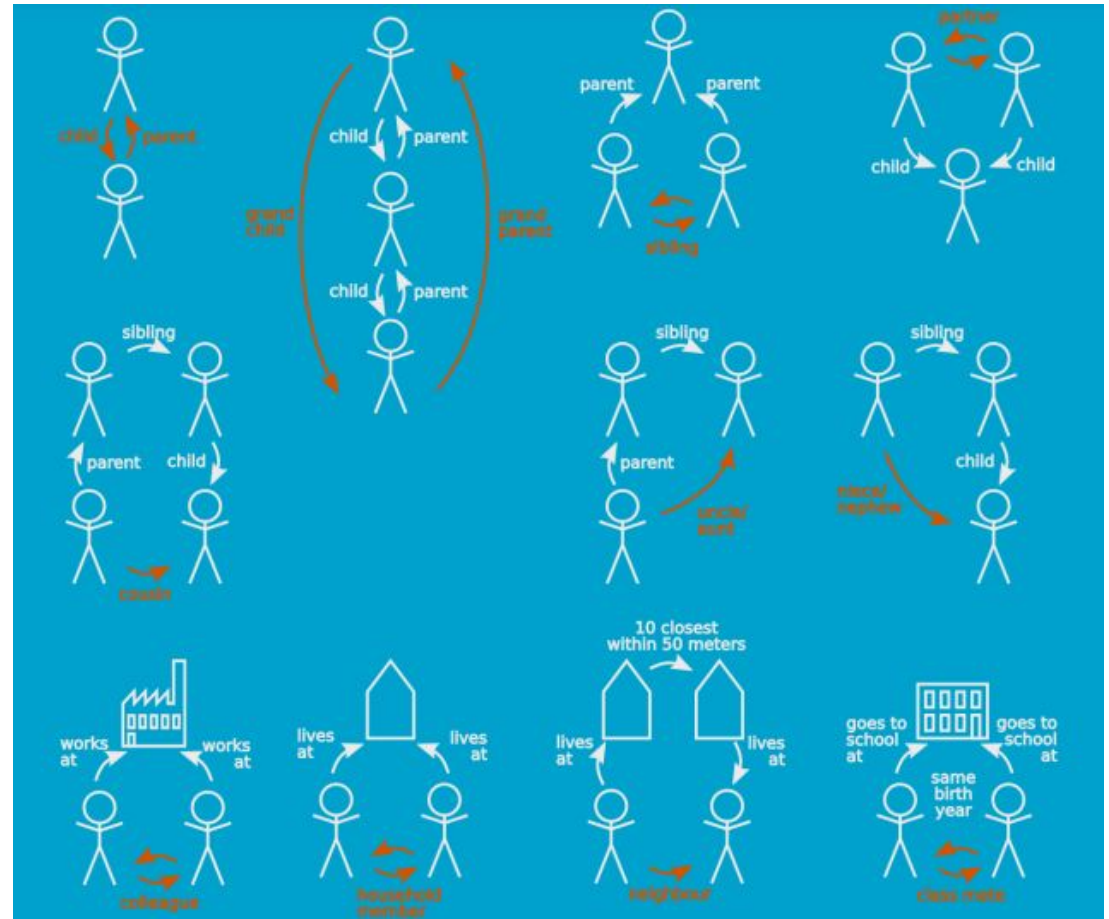


open data
challenge

More data

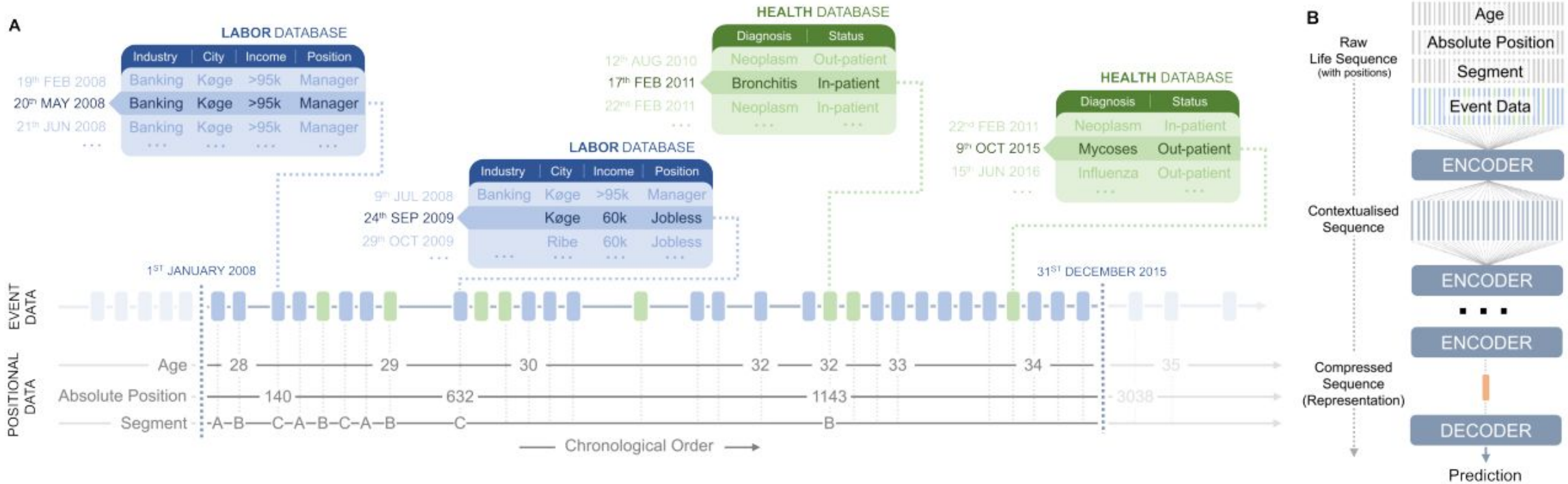
Survey + registries: access to the LISS data inside the CBS remote environment

Social network for the whole Dutch population



Picture <https://odissee-data.nl/wp-content/uploads/2019/04/PresentatieODFpilotCBS181002.pdf>

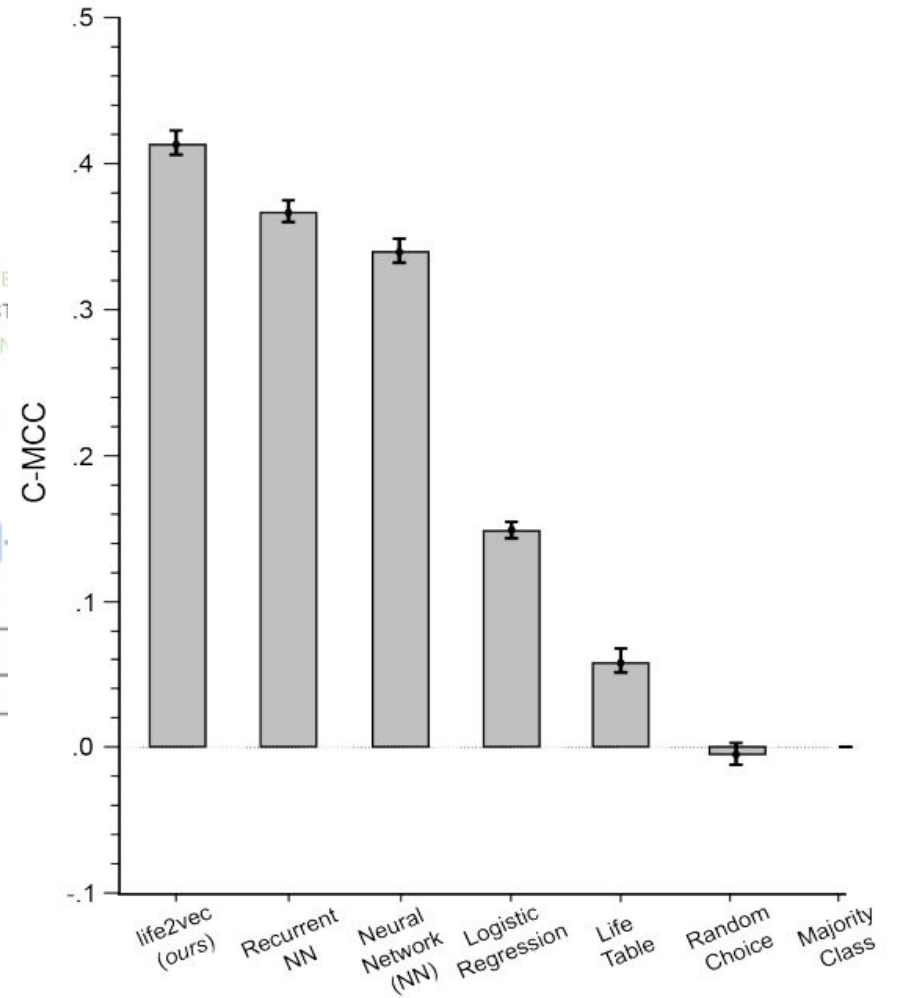
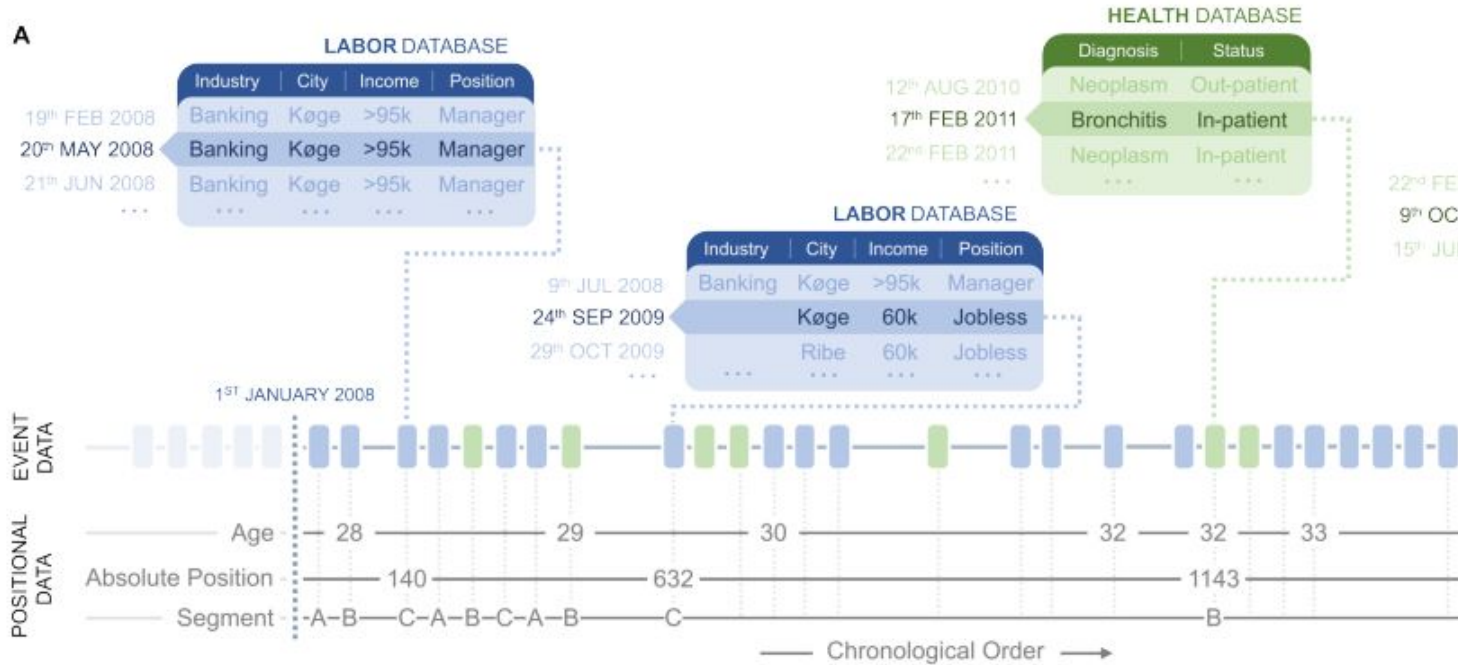
Advanced ML



Using Sequences of Life-events to Predict Human Lives

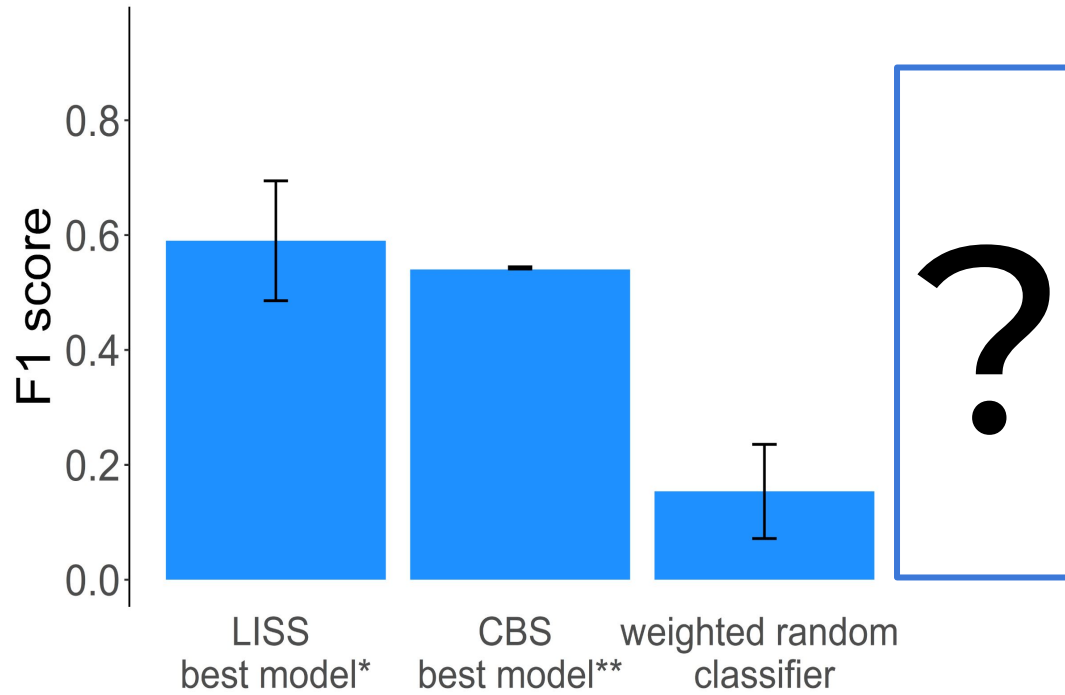
Germans Savcisens, Tina Eliassi-Rad, Lars Kai Hansen, Laust Hvas Mortensen, Lau Lilleholt, Anna Rogers, Ingo Zettler, and Sune Lehmann

Advanced ML

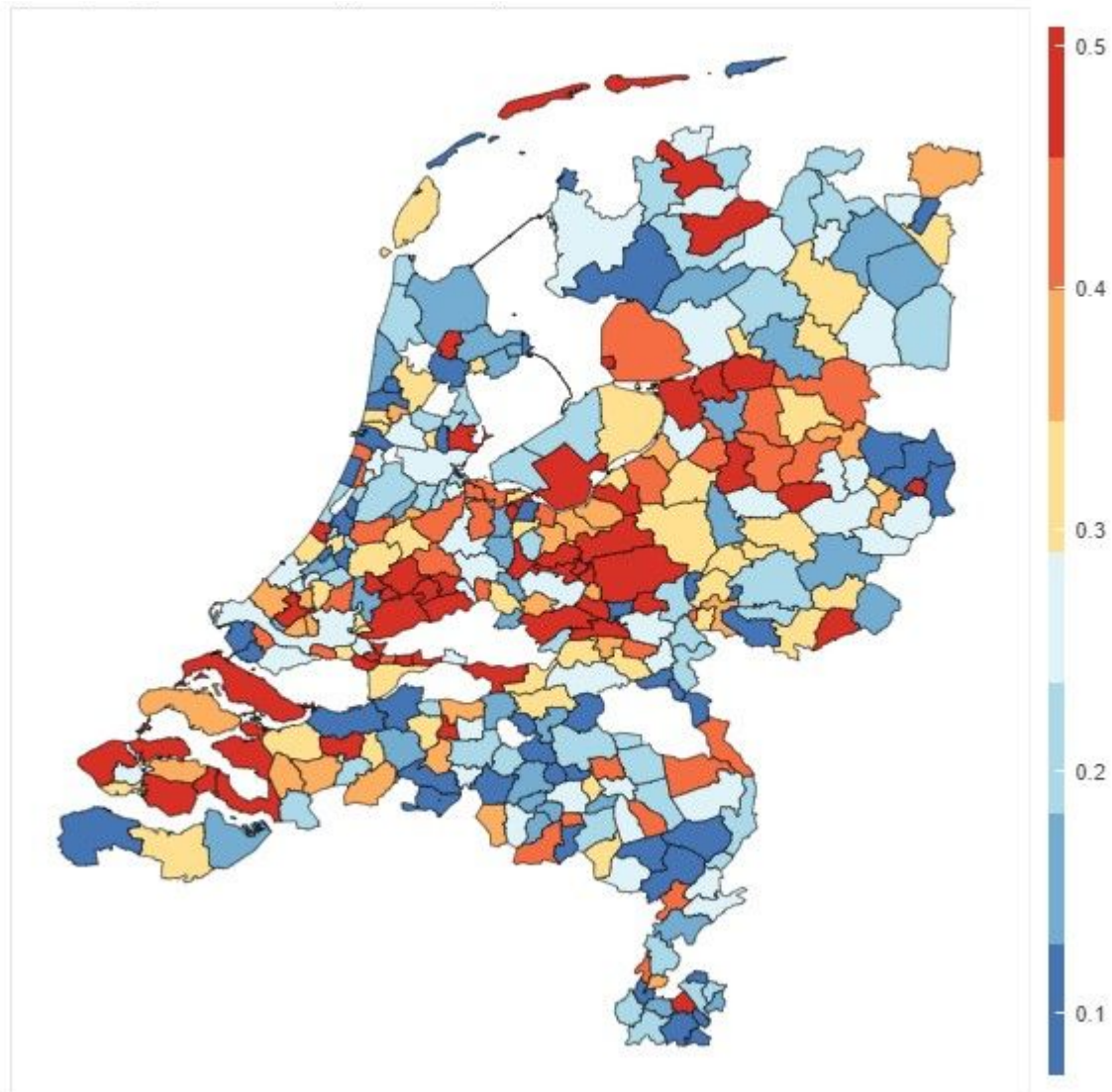


Using Sequences of Life-events to Predict Human Lives

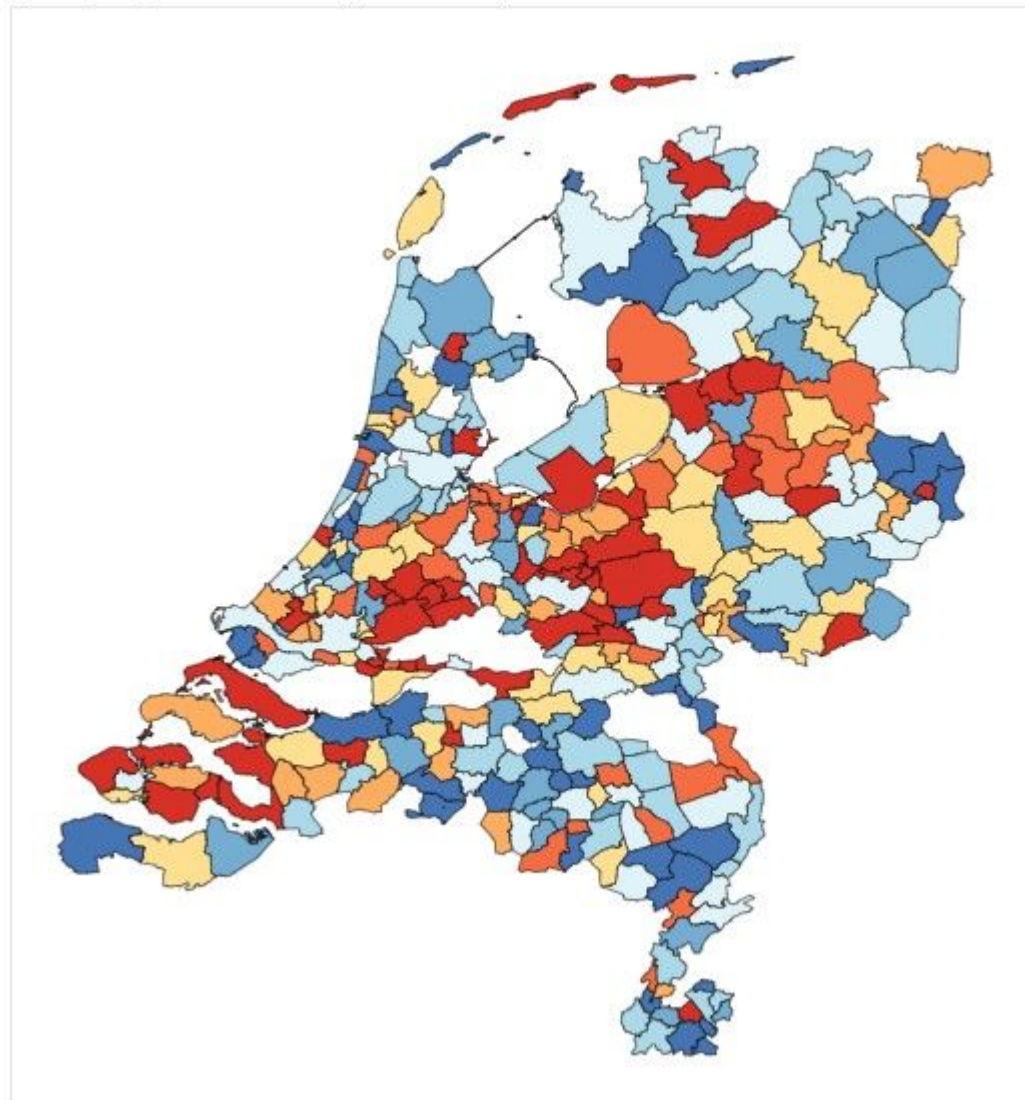
Germans Savcisens, Tina Eliassi-Rad, Lars Kai Hansen, Laust Hvas Mortensen, Lau Lilleholt, Anna Rogers, Ingo Zettler, and Sune Lehmann



- Open data challenge: assess upper limits of predictability and insights based on the error analysis, predictive performance for different groups, and comparing and interpreting the models



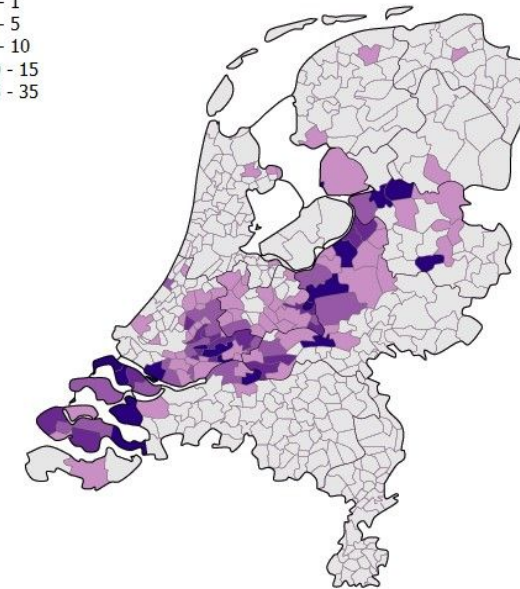
● F1 score of the predictions for people aged 18-25 in different municipalities (based on the CBS data)



© Javier Bernardo Garcia

Percentage van het aantal kiesgerechtigden

- 0 - 1
- 1 - 5
- 5 - 10
- 10 - 15
- 15 - 35



Percentage of votes for the SGP (Reformed Political Party) at the 2006 Dutch general election per municipality

https://en.wikipedia.org/wiki/Bible_Belt_%28Netherlands%29



PreFer: Data Challenge for Predicting Fertility outcomes in the Netherlands

Spring-Summer 2024

- Beat these scores
- Be a part of a unique data challenge
- Contribute to research on fertility behaviour
- Work with cool data: the LISS panel and Dutch population registries

Sign up here
to learn more:

