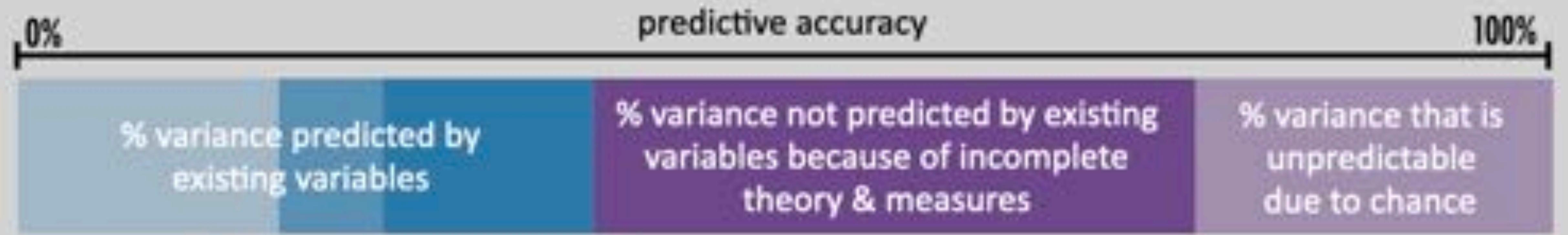


This mess we're in?

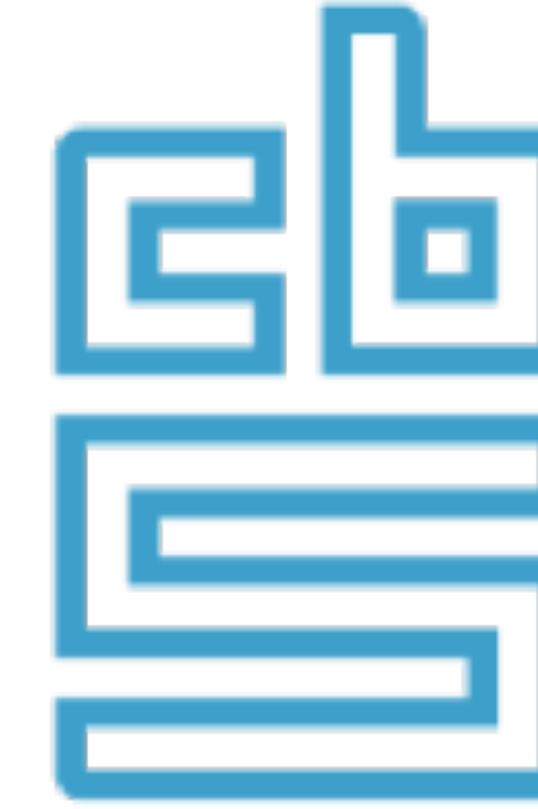
Or how simulation and prediction
will advance the social sciences



How Well Are We Doing?

variables
explain
little

Fewer births through education and flexwork?



“total effect on fertility ...
rather small

incomparable
results

Philosophical Transactions of the Royal Society B
Philosophical Transactions of the Royal Society B: Biological Sciences, Volume 368, Issue 1582, 2013, Article number 20120362, 10 pages

Health, fertility and adoptive behaviour in industrial populations

Abstract
The relationship between health and fertility has been used to explain the observed increase in the number of children born in the United States, but the relationship appears to be more complex than previously thought. This paper reviews the evidence on the relationship between health and fertility, and highlights the need for further research. The results suggest that the relationship between health and fertility is not always clear-cut, and that there may be other factors that influence fertility, such as economic conditions and social support.

Keywords
Health, fertility, adoption, birth rate, industrial populations.

Introduction
In an effort to better understand the relationship between health and fertility, it is important to consider the various factors that contribute to the observed increase in the number of children born in the United States. This paper reviews the evidence on the relationship between health and fertility, and highlights the need for further research. The results suggest that the relationship between health and fertility is not always clear-cut, and that there may be other factors that influence fertility, such as economic conditions and social support.

Conclusion
The results of this study suggest that the relationship between health and fertility is not always clear-cut, and that there may be other factors that influence fertility, such as economic conditions and social support.

Population Review
Volume 48 Number 1, 2012, pp. 1-16

From home to abroad: Fertility rates of permanent Chinese women

Abstract
Previous differences in fertility rates between Chinese women in China and abroad have been used to investigate the complex pattern of fertility in China. This paper compares the fertility rates of Chinese women in China and abroad, and explores the reasons for the difference. The results show that the fertility rates of Chinese women in China and abroad are similar, but the reasons for the difference are different. The results also show that the fertility rates of Chinese women in China and abroad are similar, but the reasons for the difference are different.

Keywords
Chinese women, fertility rates, permanent residence.

Introduction
The results of this study suggest that the relationship between health and fertility is not always clear-cut, and that there may be other factors that influence fertility, such as economic conditions and social support.

Conclusion
The results of this study suggest that the relationship between health and fertility is not always clear-cut, and that there may be other factors that influence fertility, such as economic conditions and social support.



surprising
patterns

non-replicable
results

My Upbringing in Science



PSYCHOLOGY

Estimating the reproducibility of psychological science

Open Science Collaboration*

ROYAL SOCIETY OPEN SCIENCE
rsos.royalsocietypublishing.org

The natural selection of bad science
DOI: 10.1098/rsos.160032

Res...
Ole...
276 18

Canadian Psychology / Psychologie canadienne
22(2), Vol. 46, No. 4, 211-218
<http://dx.doi.org/10.1037/can0000218>

Psychological Measurement and the Replication Crisis: Four Sacred Cows

Scott Barry Kaufman

aps

AMERICAN PSYCHOLOGICAL SOCIETY
www.apa.org/pubs/journals/psycholsci.aspx

The Theory Crisis in Psychology: How to Move Forward

Markus J. Kronen¹ and Laura F. Bringmann²
¹Department of History of Philosophy, and ²Department of Psychology and Politics, University of Konstanz



aps

False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant

Joseph P. Simmons¹, Leif D. Nelson², and Michael A. Ross³
¹University of Texas at San Antonio, ²University of California, Berkeley, and ³University of Pennsylvania

P-Curve: A Key to the File-Drawer

Tal F.D. Nelson
University of California, Berkeley

Joseph P. Simmons
University of Pennsylvania

Science

About 40% of studies in psychology are statistically significant

NEWS IN BRIEF
HOME > NEWS IN BRIEF
BIG PHARMA REPLICATES
PACIFIC STANDARD

PACIFIC STANDARD

Is there a strategy to ensure the validity of studies?

SCIENTIFIC DATA

OPEN Assessing data availability and research reproducibility in hydrology and water resources

James H. Stoggi¹, David E. Rosenberg², Adel M. Alabdullah³, Hoda Alberi⁴, Yousef A. Alabdullah⁵, & Ryan Jensen⁶

Replication (crisis) in Demography?



Reasons why not

- *Strong methods*
- *Strong focus on representative data*
- *Less measurement error*
- *Open data*
- *Large N*
- *Often descriptive*



Reasons why

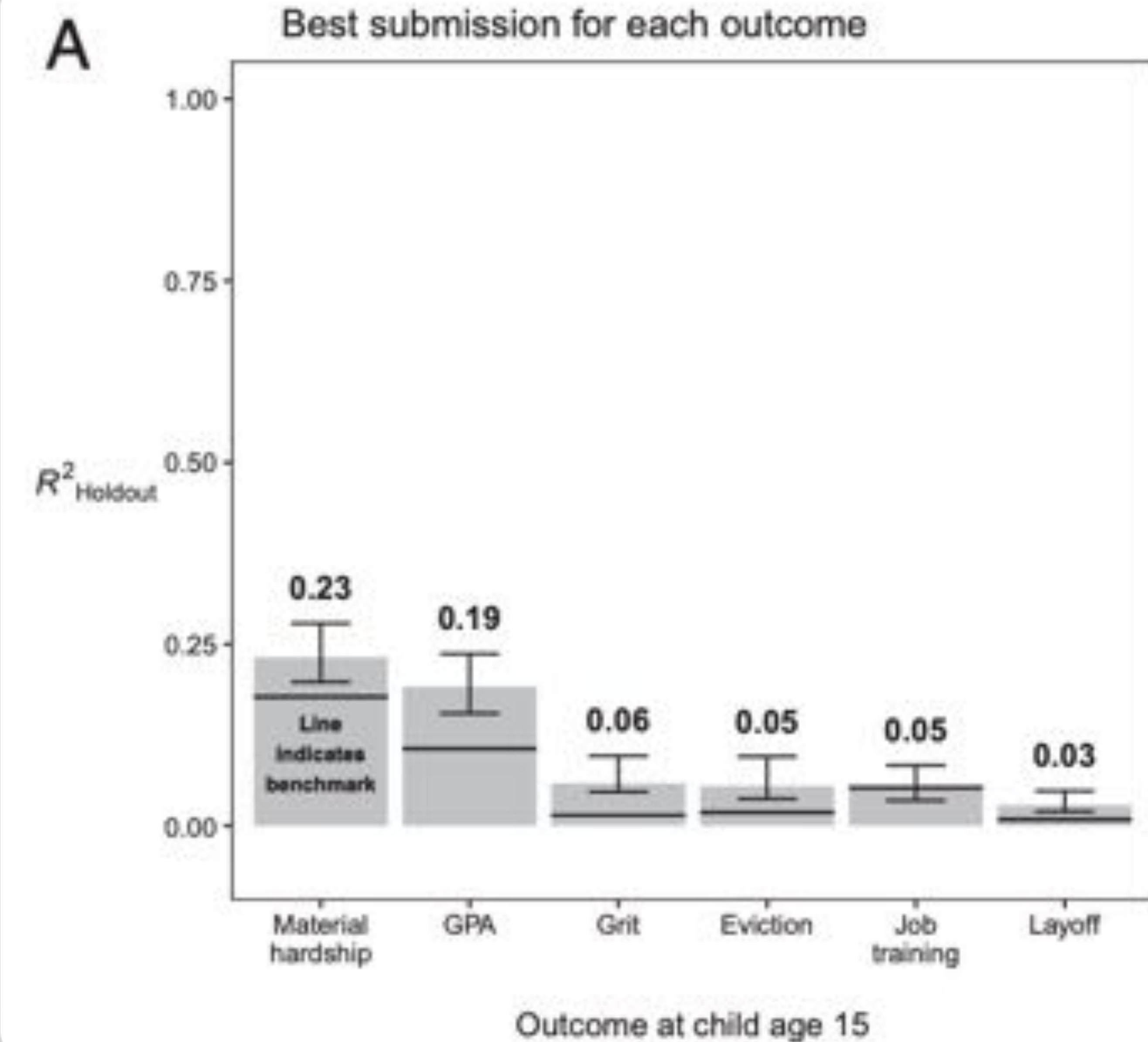
- *Non-experimental*
- *Correlational, but little causal inference*
- *Large N, yet star gazing*
- *Controlling at will*
- *“Culture” as a get-out-of-jail-for-free card*

Predictability Crisis?

Measuring the predictability of life outcomes with a scientific mass collaboration

Matthew J. Salganik^{1,2}, Ian Lundberg³, Alexander T. Kindel⁴, Caitlin E. Ahearn⁵, Khaled Al-Ghoneim⁶, Abdullah Almaatouq^{4,7}, Drew M. Altshuler⁸, Jennie E. Brand^{9,10}, Nicole Bohme Carnegie¹⁰, Ryan James Compton¹¹, Debanjan Datta¹, Thomas Davidson¹, Anna Filippova¹, Connor Gilroy¹¹, Brian J. Goode¹², Faman Jahani¹³, Ridhi Kashyap^{1,14}, Antje Kirchner¹, Stephen McKay¹, Allison C. Morgan¹⁰, Alex Pentland¹, Kivan Polimis¹, Louis Raes¹⁰, Daniel E. Rigobon¹, Claudia V. Roberts⁷, Diana M. Stanescu¹, Yoshihiko Suhara¹⁴, Adnan Usmani¹⁵, Erik H. Wang¹, Muna Adem¹⁶, Abdulla Alhajri¹⁷, Bedoor AlShebli¹⁸, Redwane Amin¹⁹, Ryan B. Amos¹, Lisa R. Argyle¹⁰, Livia Baer-Bositis²⁰, Moritz Büchi¹⁹, Bo-Ryehn Chung¹, William Eggert¹, Gregory Faletto¹¹, ZhiLin Fan¹, Jeremy Freese²¹, Tejomay Gadgil¹⁹, Josh Gagne¹⁹, Yue Gao¹⁹, Andrew Halpern-Manners¹⁶, Sonia P. Hashmi¹, Sonia Hausen²², Guanhua He²², Kimberly Higuera¹⁹, Bernie Hogan²², Ilana M. Horwitz²², Lisa M. Hummel²², Naman Jain¹, Kun Jin¹⁰, David Jurgens¹², Patrick Kaminski^{16,22}, Areg Karapetyan^{22,23}, E. H. Kim²¹, Ben Leizman¹, Naija Liu¹, Malte Möser¹, Andrew E. Mack¹, Mayank Mahajan⁷, Noah Mandell²⁴, Helge Marahrens²², Diana Mercado-Garcia²², Viola Mocz²², Katarina Mueller-Gantell²², Ahmed Musse²², Qiankun Niu²², William Nowak¹, Hamidreza Omidvar²², Andrew Orr¹, Karen Ouyang¹, Katy M. Pinto²², Ethan Porter²², Kristin E. Porter²², Crystal Qian⁷, Tamkinat Rau²², Anahit Sargsyan²², Thomas Schaffner¹, Landon Schnabel²², Bryan Schonfeld¹, Ben Sender²², Jonathan D. Tang¹, Emma Turkov²², Austin van Loon²², Onur Varol^{22,23}, Xiafei Wang²², Zhi Wang^{22,23}, Julia Wang²², Flora Wang²², Samantha Weissman²², Kirstie Whitaker^{22,23}, Maria K. Wolters^{22,23}, Wei Lee Woon²², James Wu^{22,23}, Catherine Wu²², Kengran Yang²², Jingwen Yin²², Bingyu Zhao²², Chenyun Zhu²², Jeanne Brooks-Gunn^{22,23}, Barbara E. Engelhardt²², Moritz Hardt²², Dean Knox¹, Karen Levy²², Arvind Narayanan⁷, Brandon M. Stewart¹, Duncan J. Watts^{22,23,24}, and Sara McLanahan¹.

data challenge:
predicting life outcomes
based on ~6000 variables
by 160 teams
both theory- & data-driven



Predictability Crisis?

“

Social scientists studying the life course must find a way to reconcile a widespread belief that understanding has been generated by these data—as demonstrated by more than 750 published journal articles using the Fragile Families data with the fact that the very same data could not yield accurate predictions of these important outcomes.

How Well Are We Doing?

The Proposal

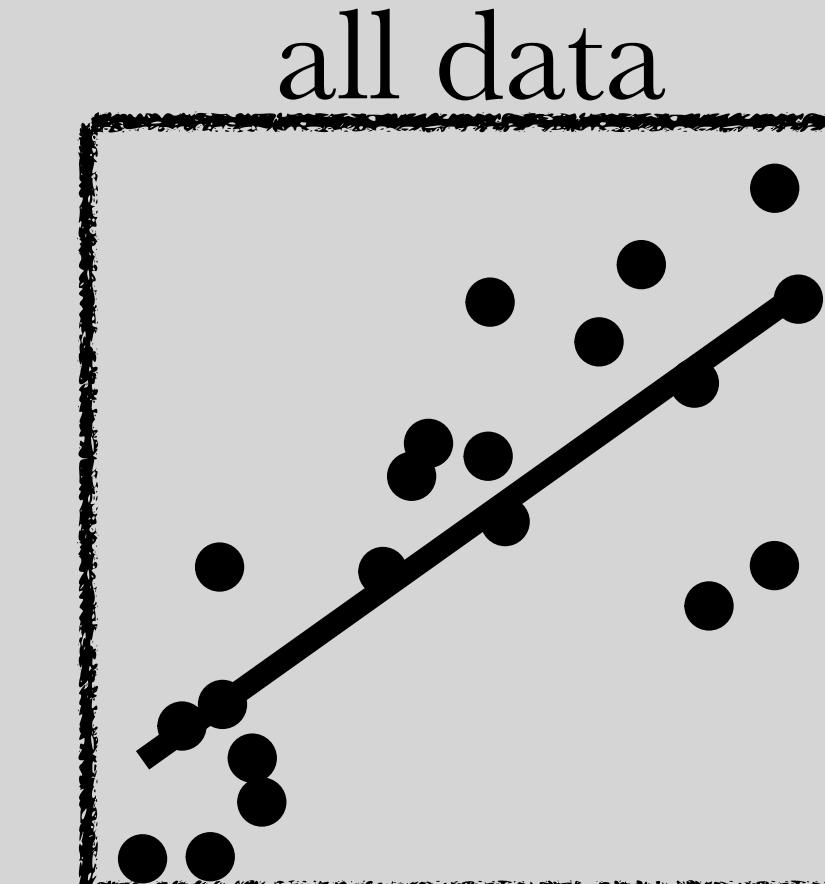
a shift towards **prediction**
leads to a more reliable
and useful social science

microsimulation can
advance traditional
statistical modelling

Take-Home Messages

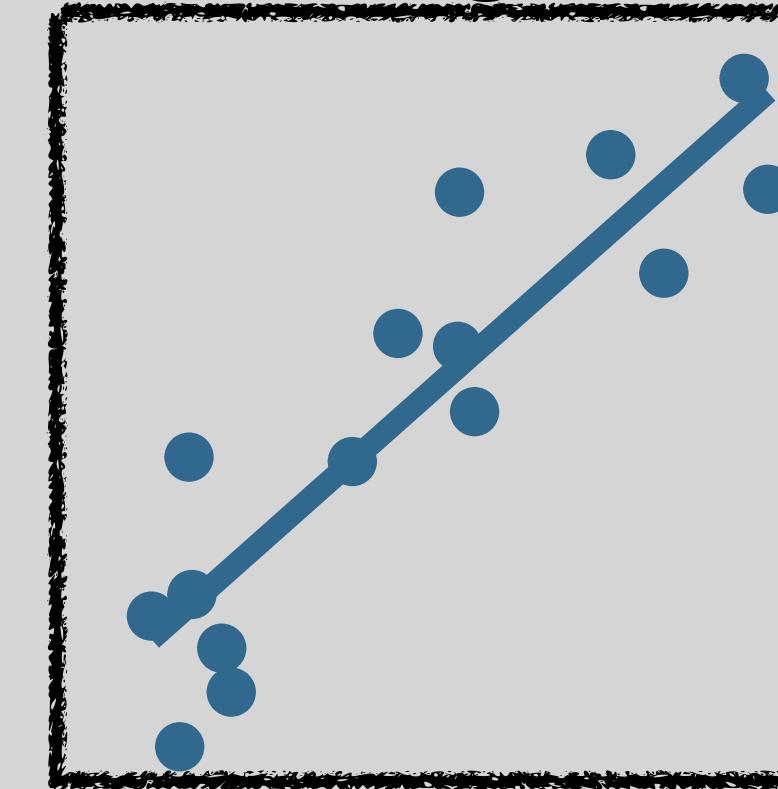
a shift towards **prediction**
leads to a more reliable
and useful social science

Out-of-Sample Prediction



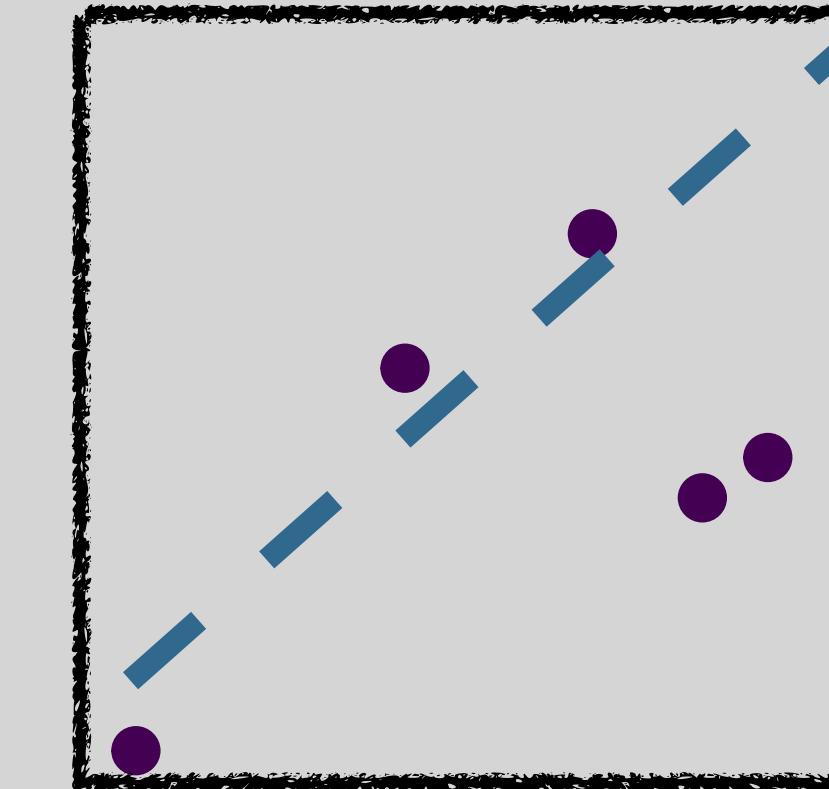
RMSE: 0.41

training data



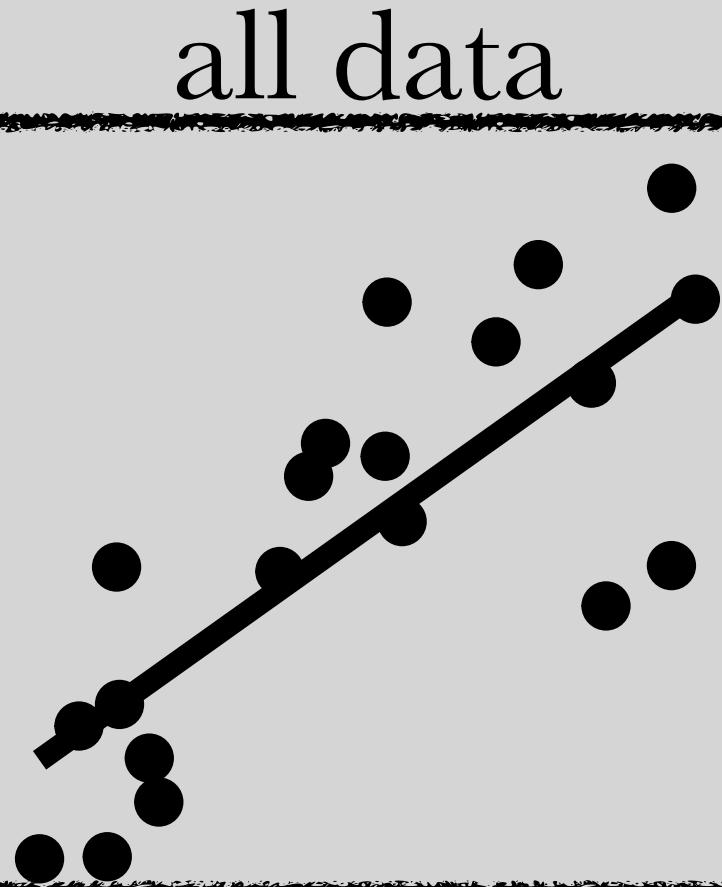
RMSE: 0.41

test data



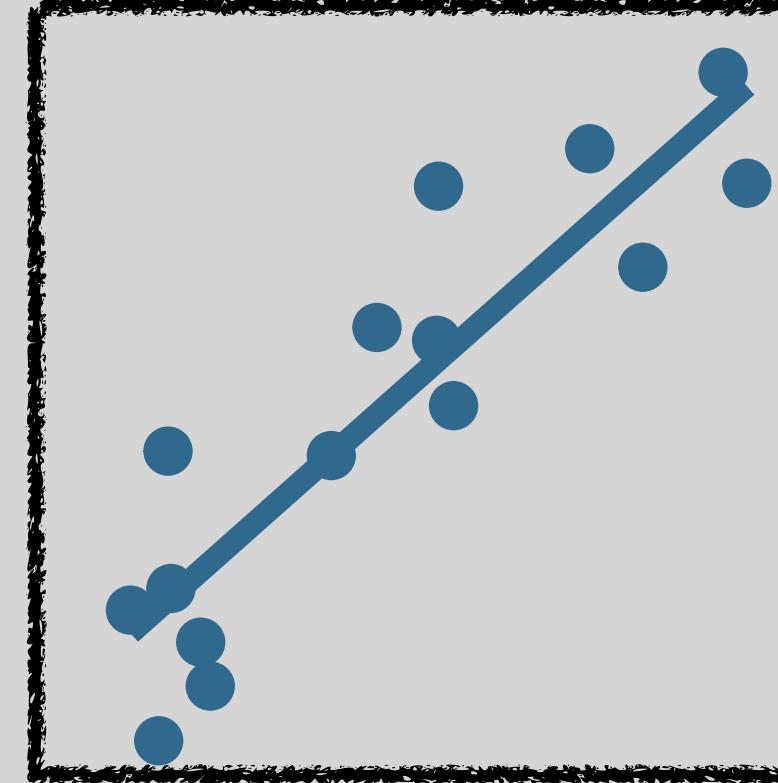
RMSE: 0.45

Out-of-Sample Prediction



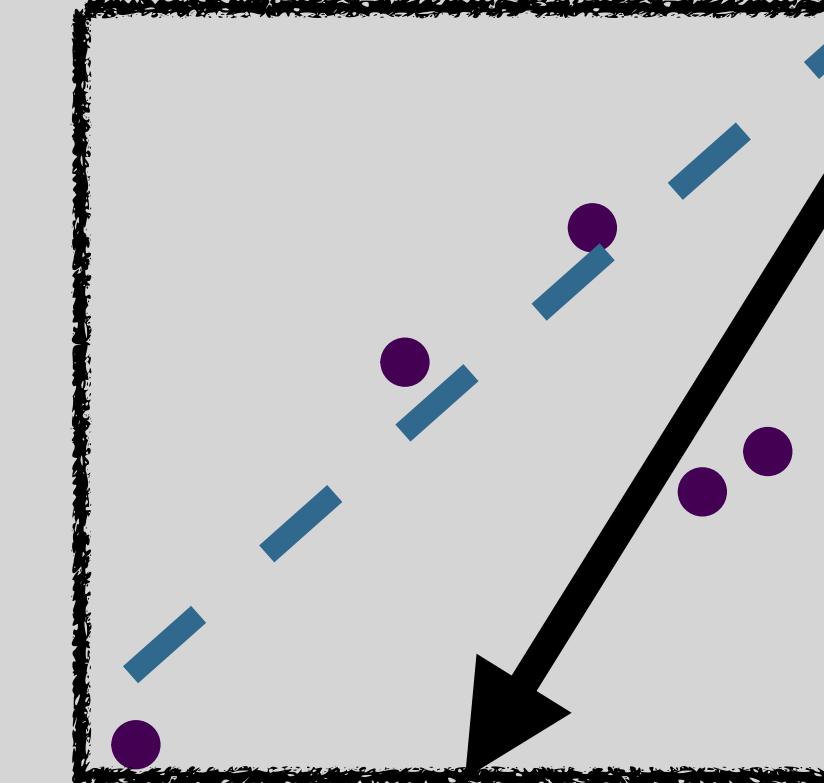
RMSE: 0.41

training data



RMSE: 0.41

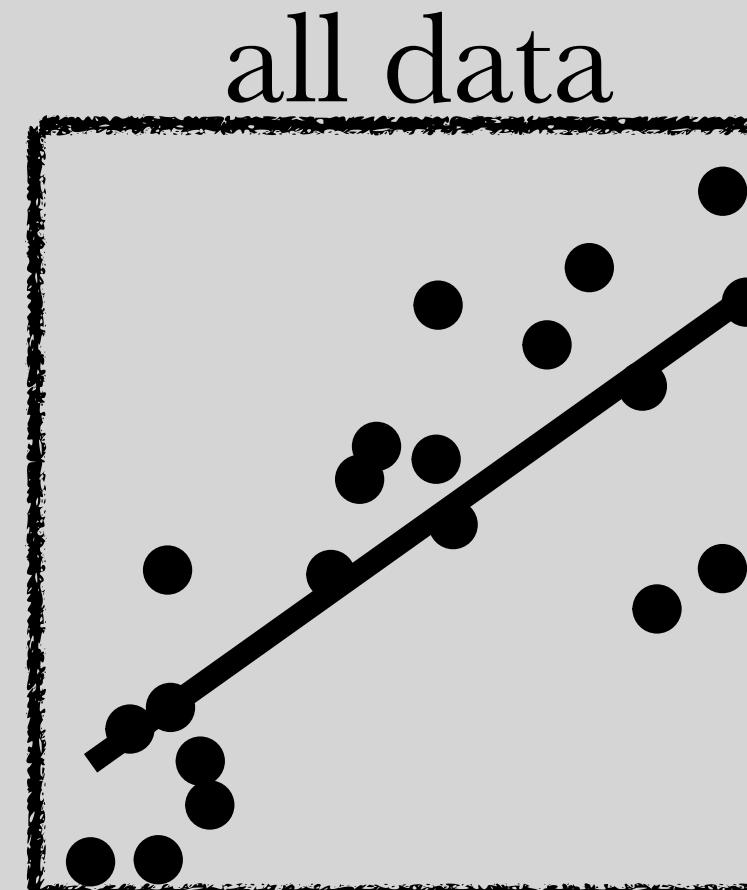
test data



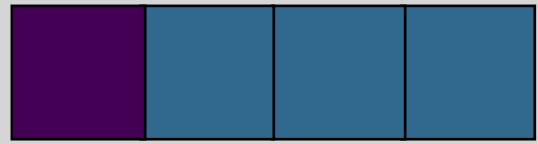
RMSE: 0.45

out-of-sample
predictive ability

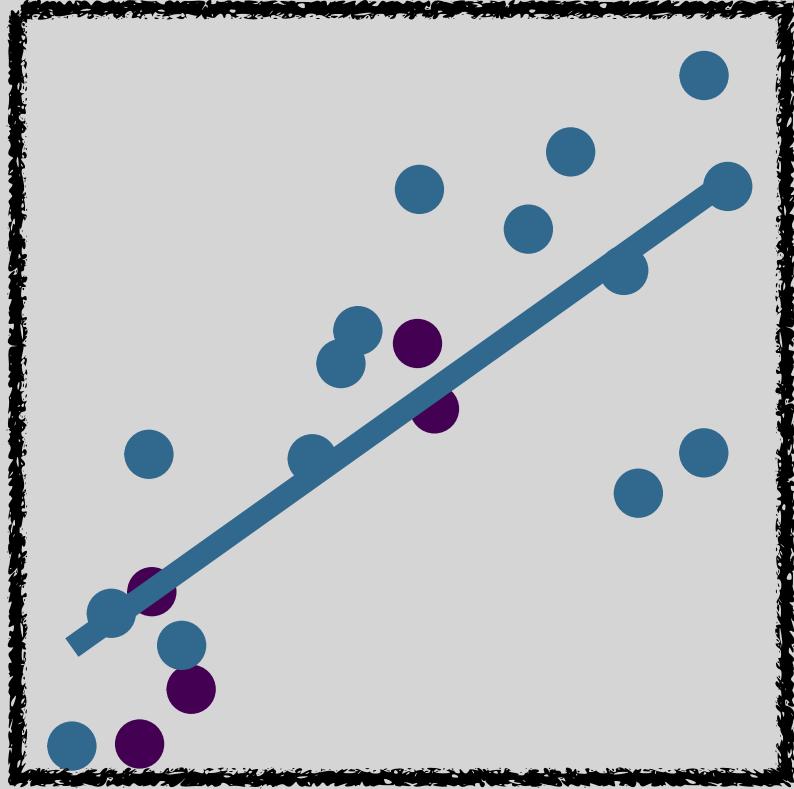
Cross-Validation



RMSE: 0.41



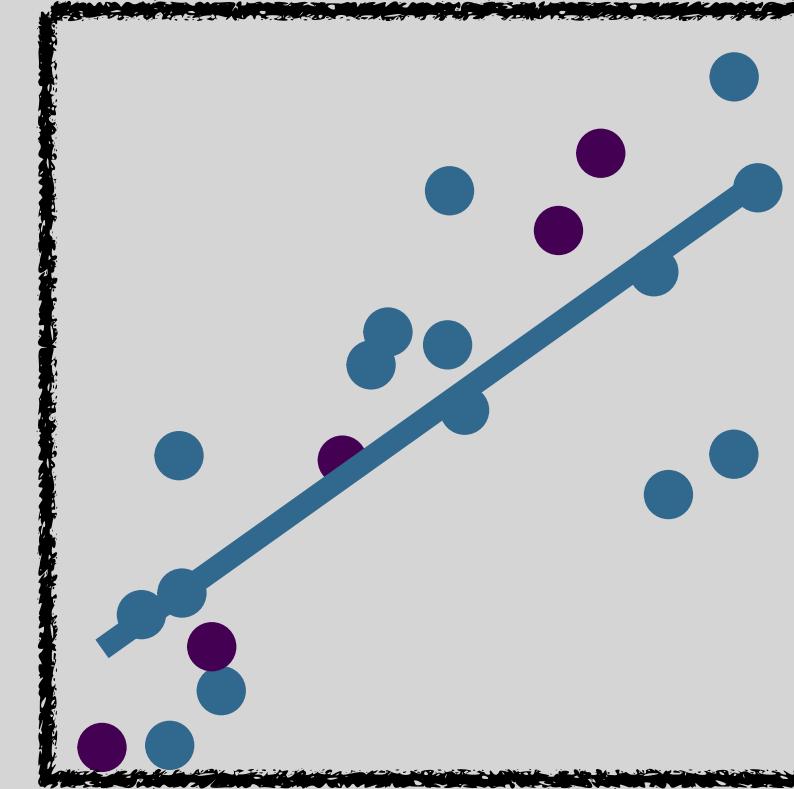
fold 1



RMSE: 0.38



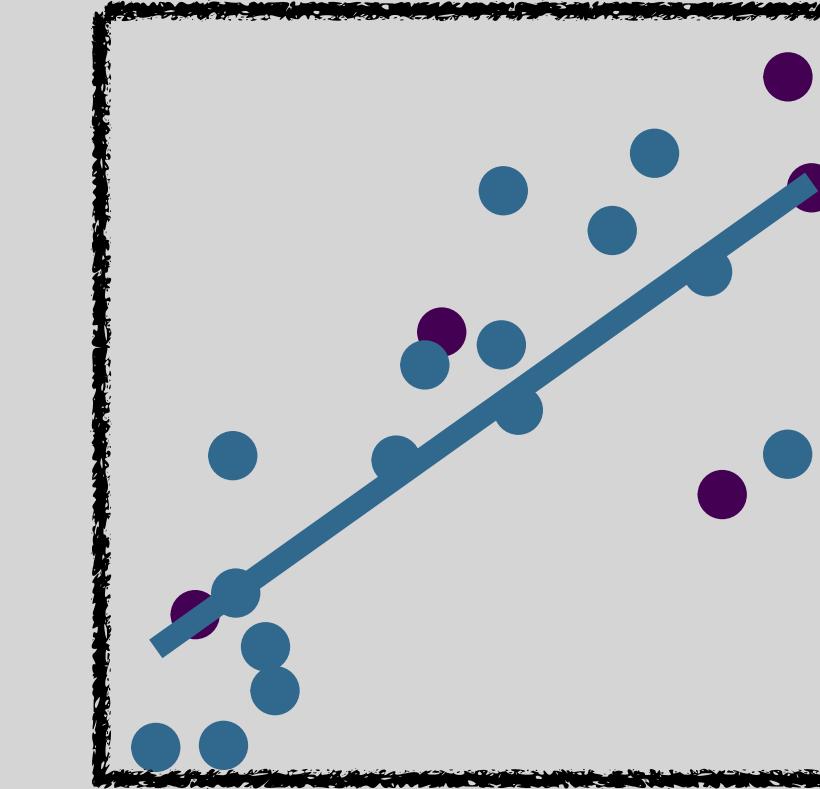
fold 2



RMSE: 0.38



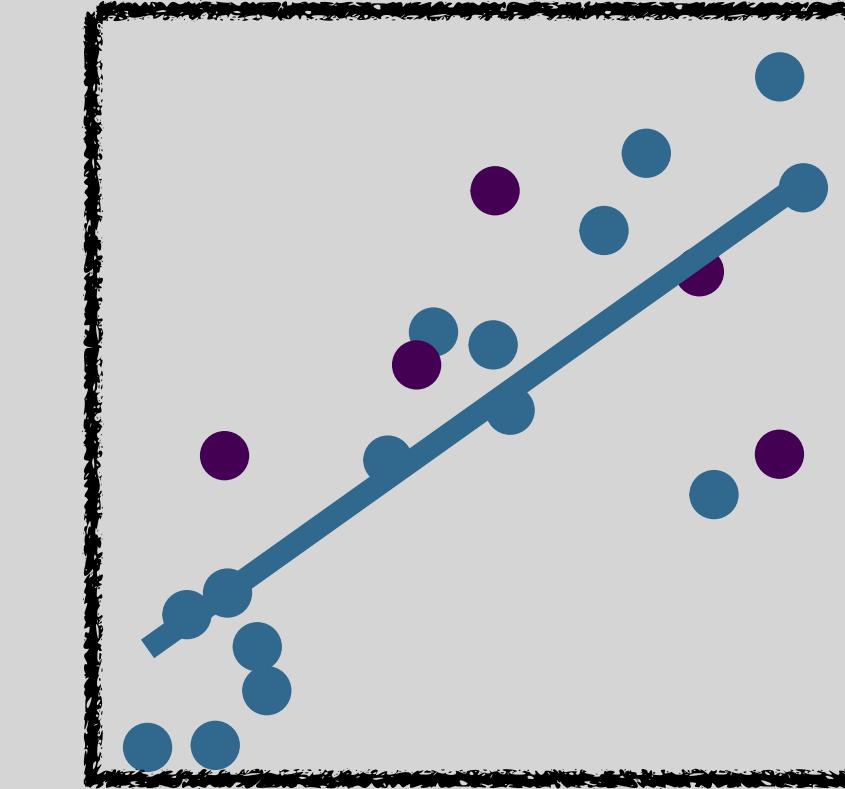
fold 3



RMSE: 0.45

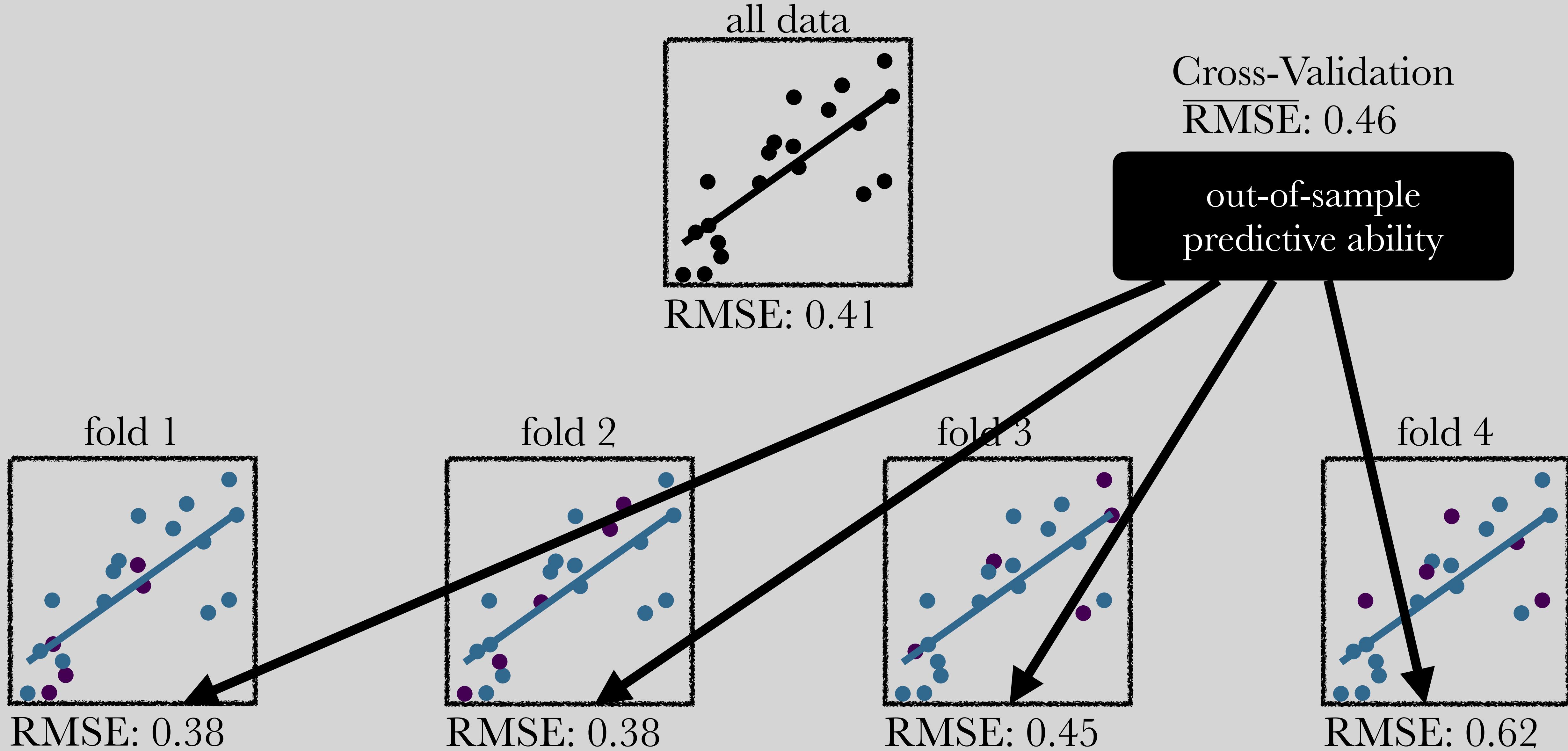


fold 4



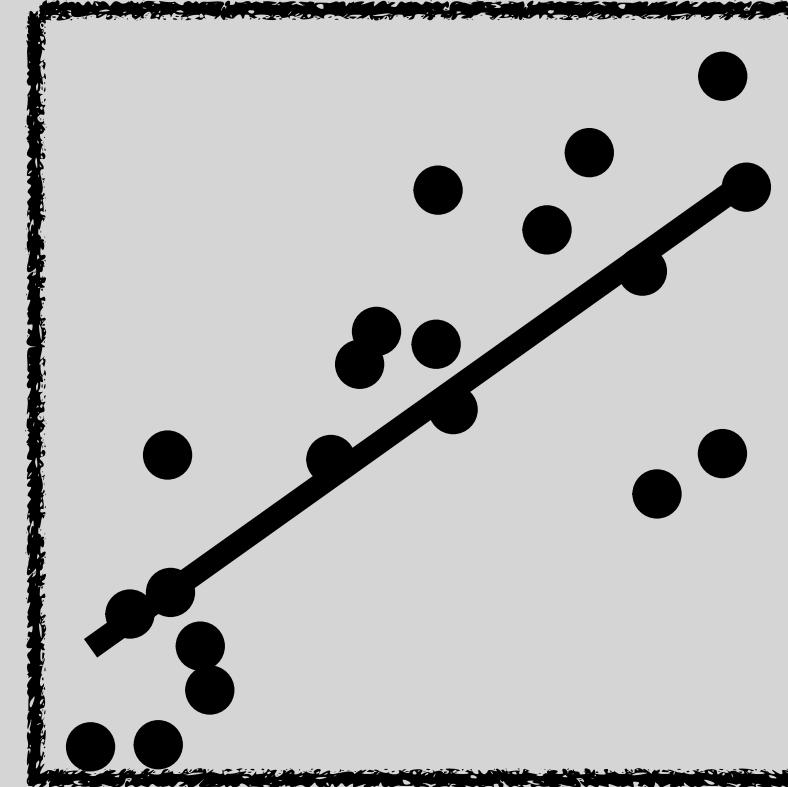
RMSE: 0.62

Out-of-Sample Prediction



Out-of-Sample Prediction

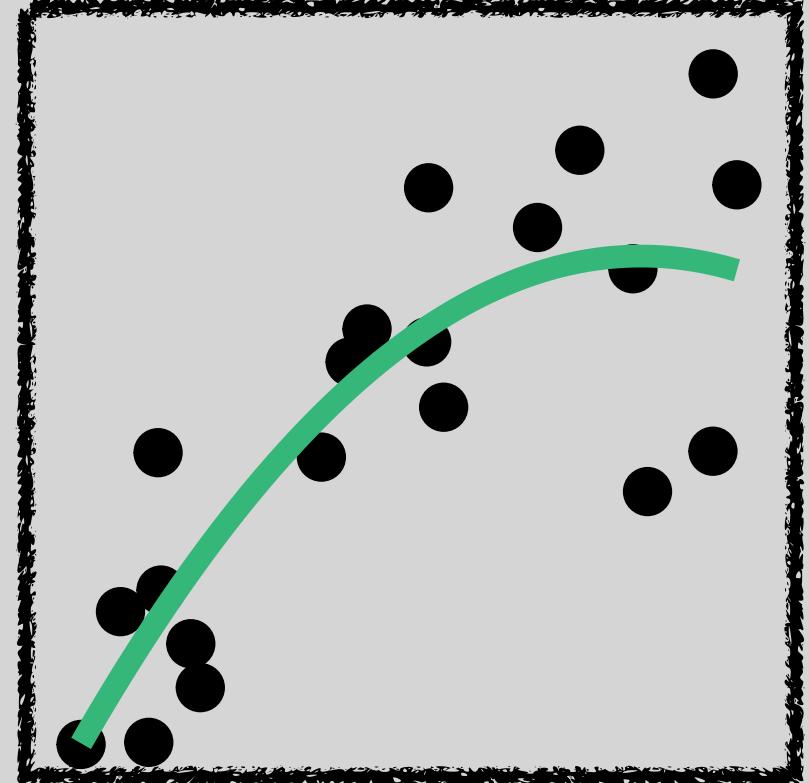
all data



RMSE: 0.41

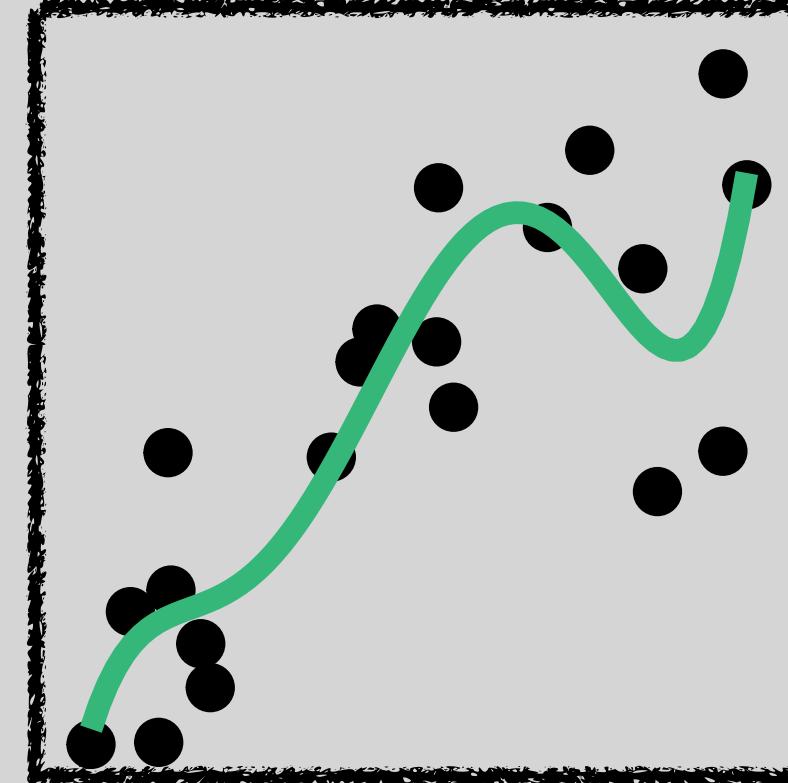
CV RMSE: 0.46

2 polynomials



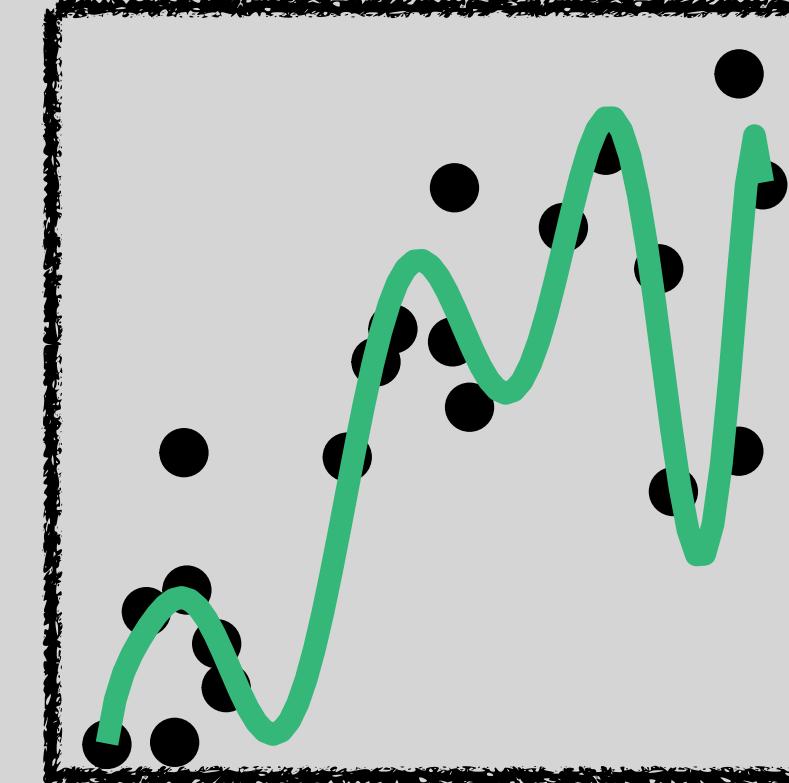
RMSE: 0.36

5 polynomials



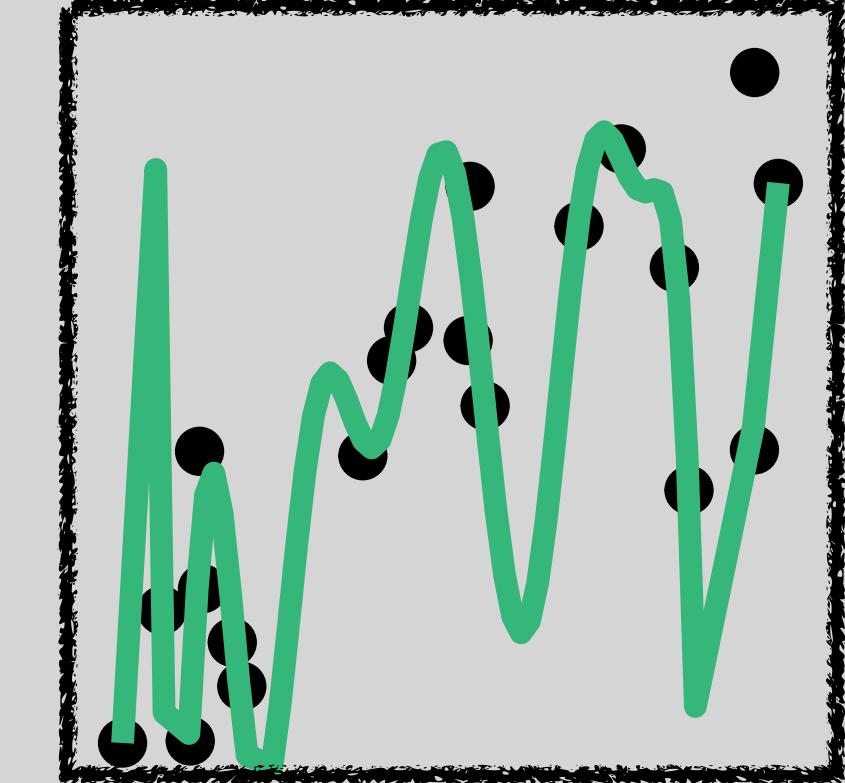
RMSE: 0.33

10 polynomials



RMSE: 0.28

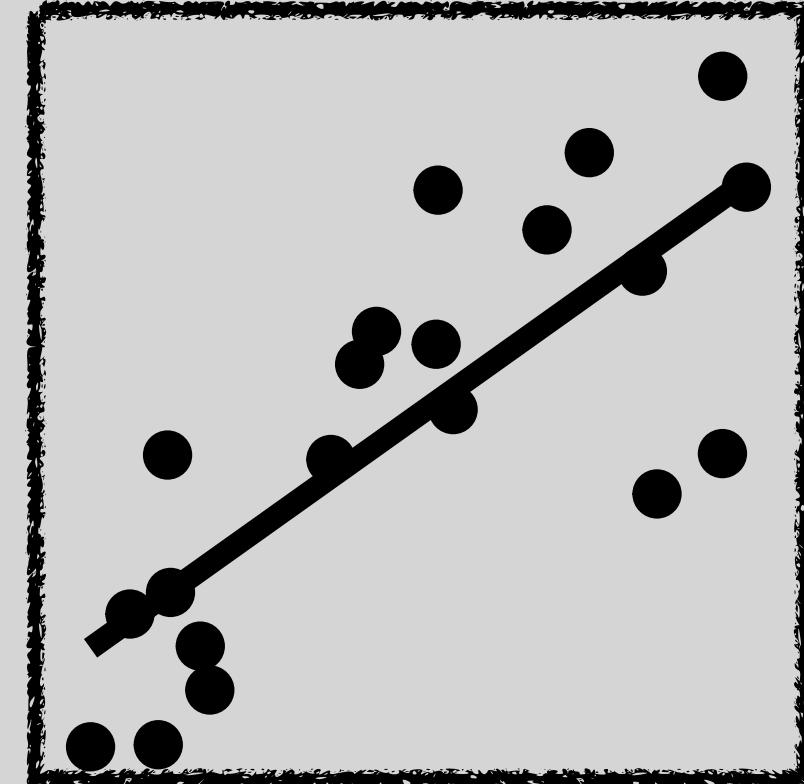
14 polynomials



RMSE: 0.22

Out-of-Sample Prediction

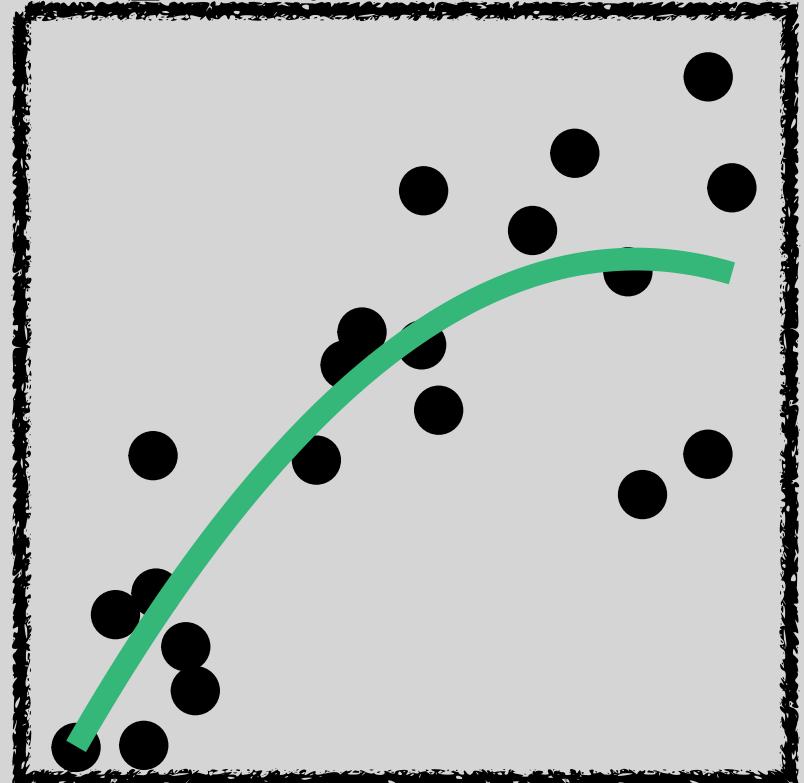
all data



RMSE: 0.41

CV RMSE: 0.46

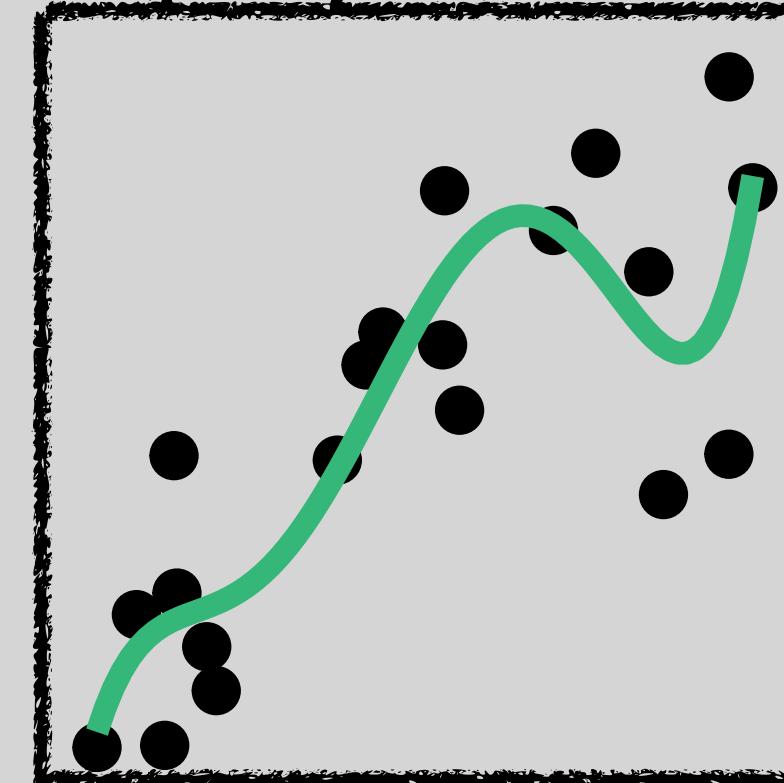
2 polynomials



RMSE: 0.36

CV RMSE: 0.44

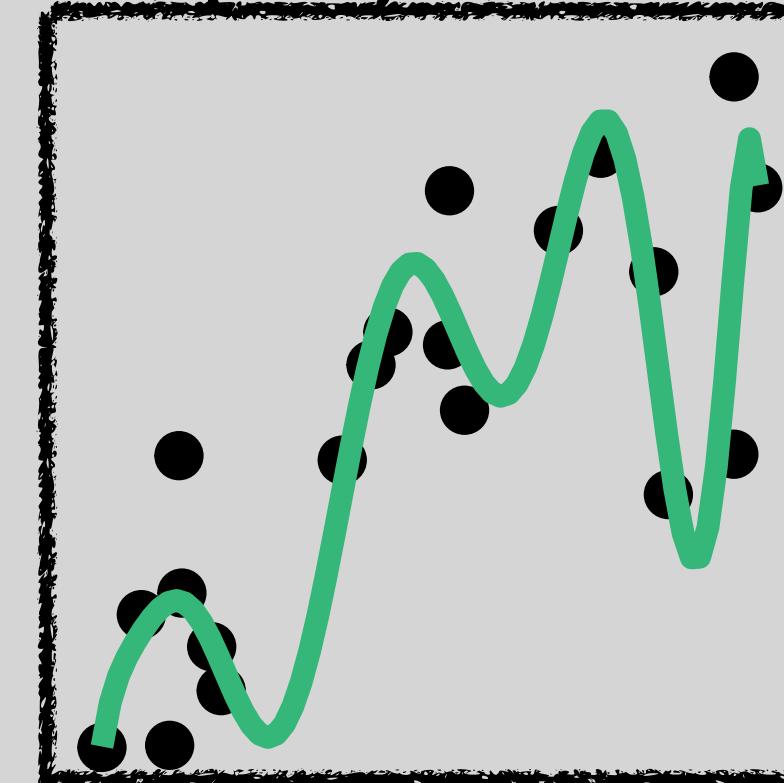
5 polynomials



RMSE: 0.33

CV RMSE: 0.58

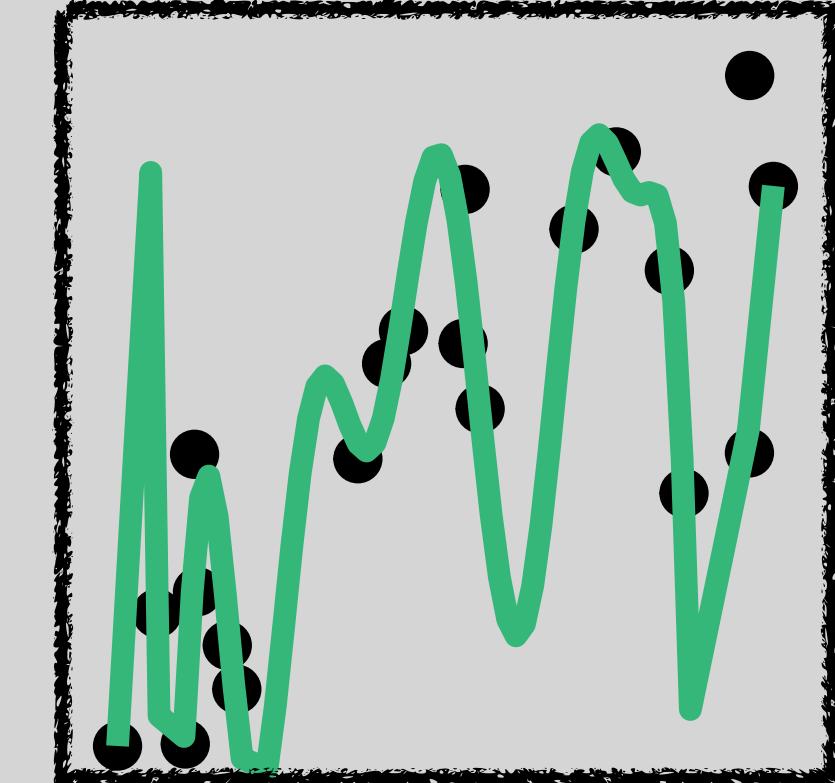
10 polynomials



RMSE: 0.28

CV RMSE: 6.4

14 polynomials

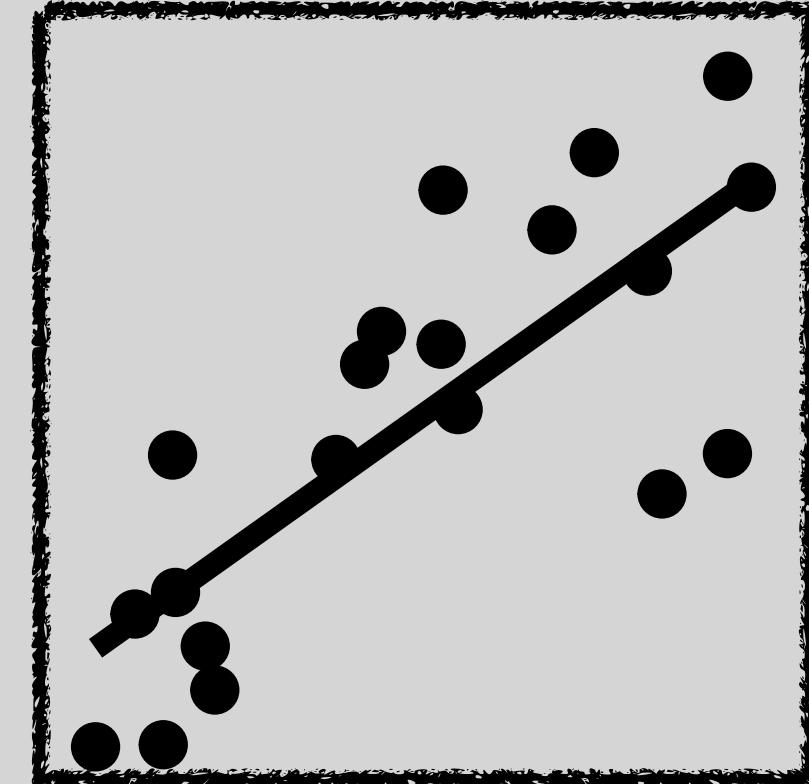


RMSE: 0.22

CV RMSE: 1206

Out-of-Sample Prediction

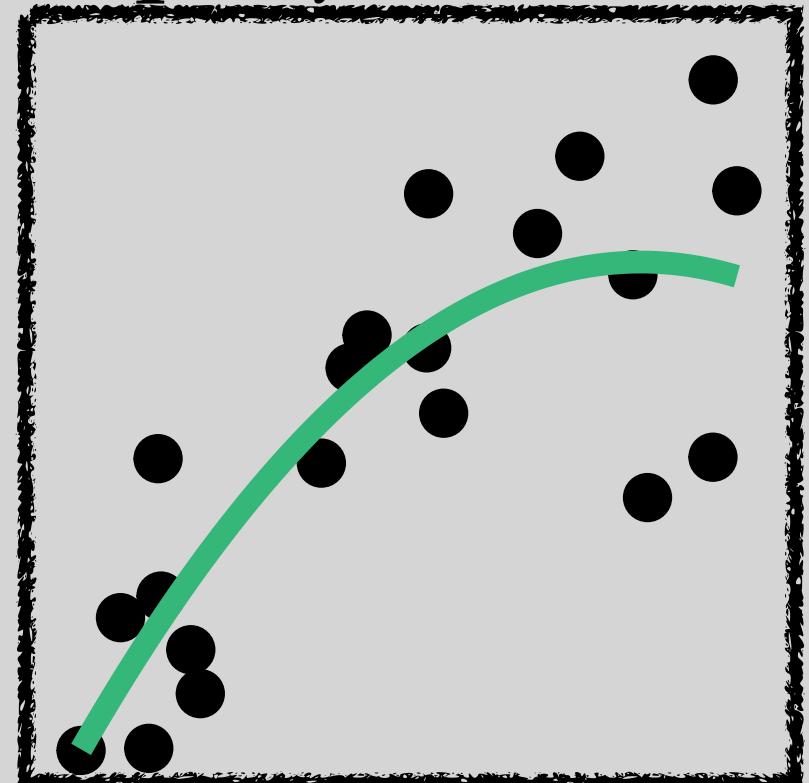
all data



RMSE: 0.41

CV RMSE: 0.46

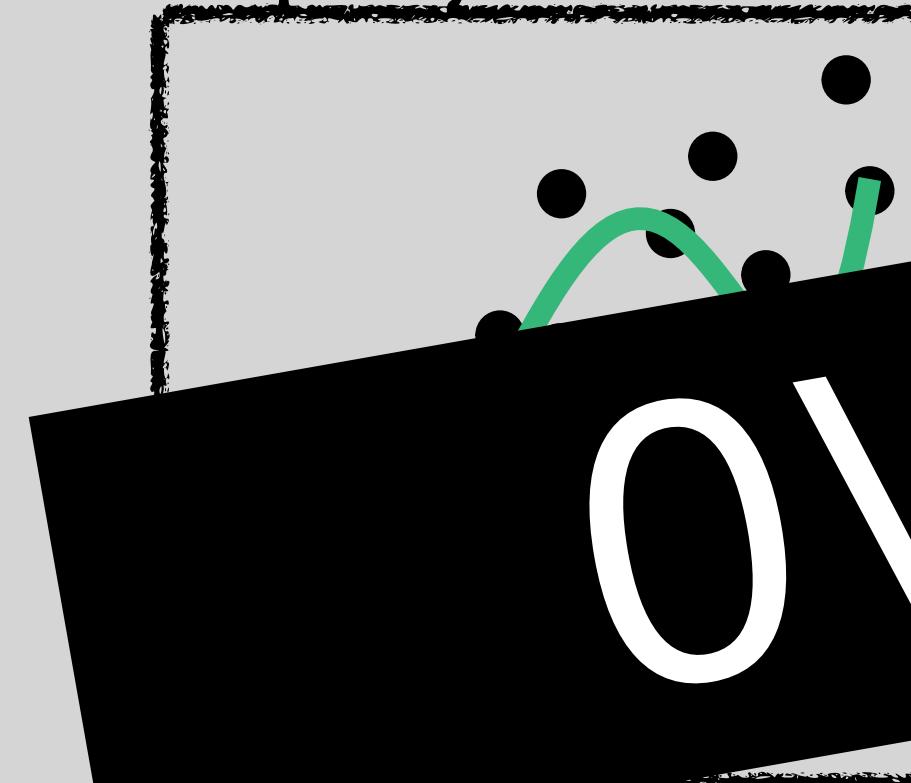
2 polynomials



RMSE: 0.36

CV RMSE: 0.44

5 polynomials



RMSE: 0.33

CV RMSE: 0.58

10 polynomials



RMSE: 0.28

CV RMSE: 6.4

OVERFITTING



RMSE: 0.22

CV RMSE: 1206

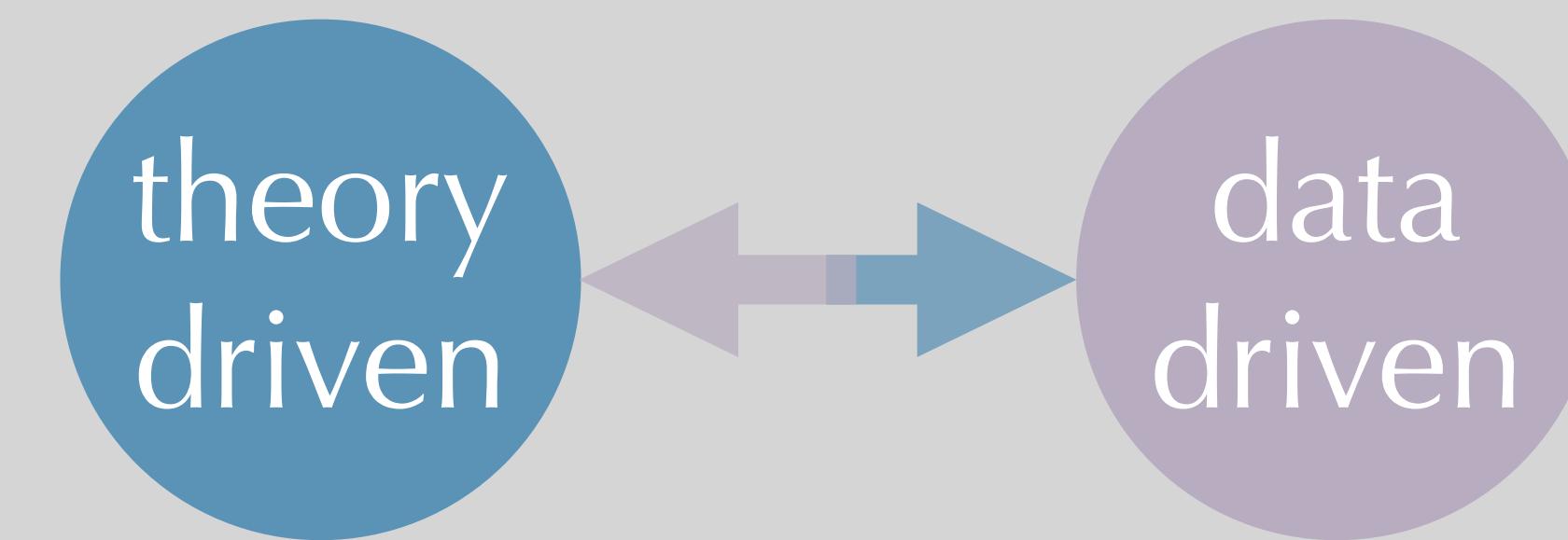
The Proposal

a shift towards **prediction**
leads to a more reliable
and useful social science

out-of-sample predictive ability:



clear measure of
effect size



facilitates dialogue
theory- and data-
driven models



measure of distance
theory and practice



out-of-sample predictive ability

- *is easy(ier) to understand*
- *can be compared across analytical techniques*
- *can be compared across models*
- *is less gameable*

Barriers to Training in: Review, November 19, 2008, 33-47
<http://dx.doi.org/10.1146/annurev-statistica-060507-090747>

Logistic Regression: Why We Cannot Do What We Think We Can Do, and What We Can Do About It

Craig Head

Logistic regression estimates do not reflect the linear or important aspect. They are affected by variables associated with the dependent variable in the model. This often has great largely consistent by nonlogit, important, important logit terms or other terms in other models. The degree of nonlinearity is the model. Nonlogit terms are with other the other models are given or other models with different independent variables in those problems and possible ways of estimating them.

Introduction

The use of logistic regression is routine in the social sciences when studying outcome that are binary or discrete represented by true/false, simple or many in classification research (education levels, gender, ethnicity, marital status, etc.). However, logistic regression is not without its own set of challenges. In this paper, we will discuss some of these challenges, including interpretation of coefficients, odds ratios, and much more. We will also discuss how to interpret odds ratios in logistic regression when there is a dichotomous dependent variable, categorical independent variables, and logistic regression and the problem of potentially inconsistent in methods in parameter estimation. However, we examine ways of interpreting odds ratios, logistic regression, have some important problems.

The problem arise from noninterpretability of the logit and the logit odds ratio as a model of variables that affect an outcome. We will also discuss how to interpret odds ratios in logistic regression when there is a continuous dependent variable.

© The Author 2008. Published by Blackwell Publishing Ltd. All rights reserved.
For permission, please e-mail: permissions@oxfordjournals.org



Annual Review of Sociology
Interpreting and Understanding
Logits, Probits, and Other
Nonlinear Probability Models

Richard Green,¹ Kristian Bjørnstad Karlsøn,²
and Anders Hahn³

¹University College Dublin, Dublin, Ireland; ²University of Oslo, Oslo, Norway;
³Department of Sociology, University of Wisconsin-Madison, Madison, WI, USA



Volume 37 Number 1 March 2008
ISSN 0094-9711
10.1111/j.1540-6598.2008.00671.x

The Annual Review of Sociology is a peer-reviewed journal.

Copyright © 2008 by Blackwell Publishing Ltd.

Published online in Wiley InterScience.

Keywords

Logit, probit, GLIM method, standardized, marginal effects, linear probability, conditional logistic.

Abstract

Methodologists in sociology and other social sciences constantly encounter the use of the logit or probit function to measure variables affecting an outcome. Logit or probit regression is a standard, and analogous to logit when there are more than two categories. But these methods always provide only odds ratios or the inverse of a logit coefficient that, over the past 10 years, has proved to publications experts of their nonlinear probability models and, particularly, in attempts to interpret their parameters. In this review, we focus on the literature to explain the problems, show how they must be addressed to correctly discuss the strengths and weaknesses of decisions that are being suggested, and point to lines of further analysis.

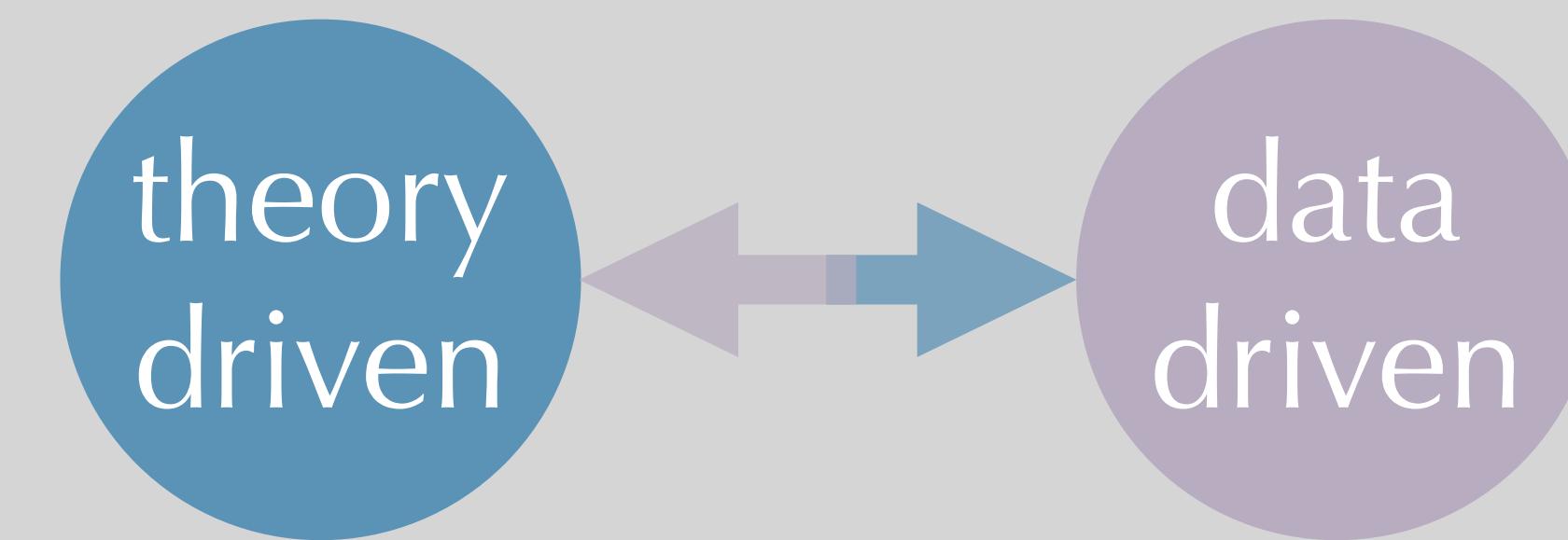
The Proposal

a shift towards **prediction**
leads to a more reliable
and useful social science

out-of-sample predictive ability:



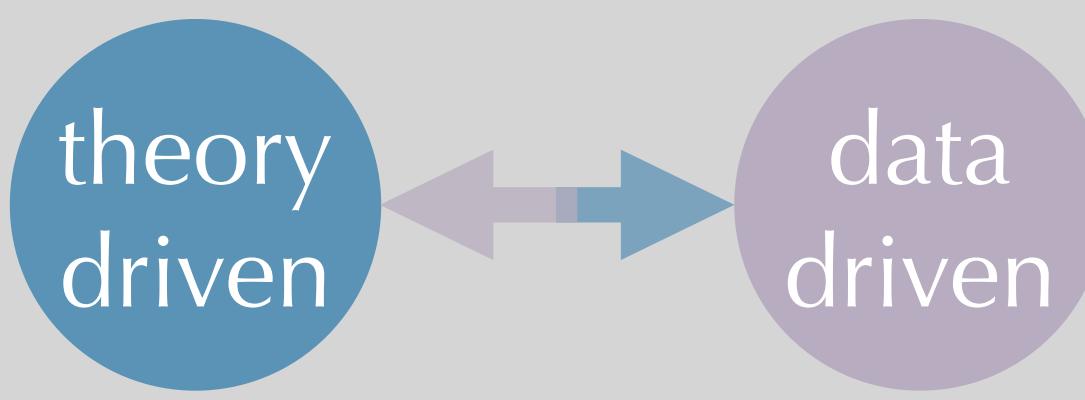
clear measure of
effect size



facilitates dialogue
theory- and data-
driven models



measure of distance
theory and practice



theory-driven *vs* data-driven

focus on (causal) estimates

support based on p-value

limited number of variables (k)

NHST weird theory-testing

long reign the linear model

pet variable problem

focus on predictive ability

support based on prediction

k may be larger than n



estimates less interpretable

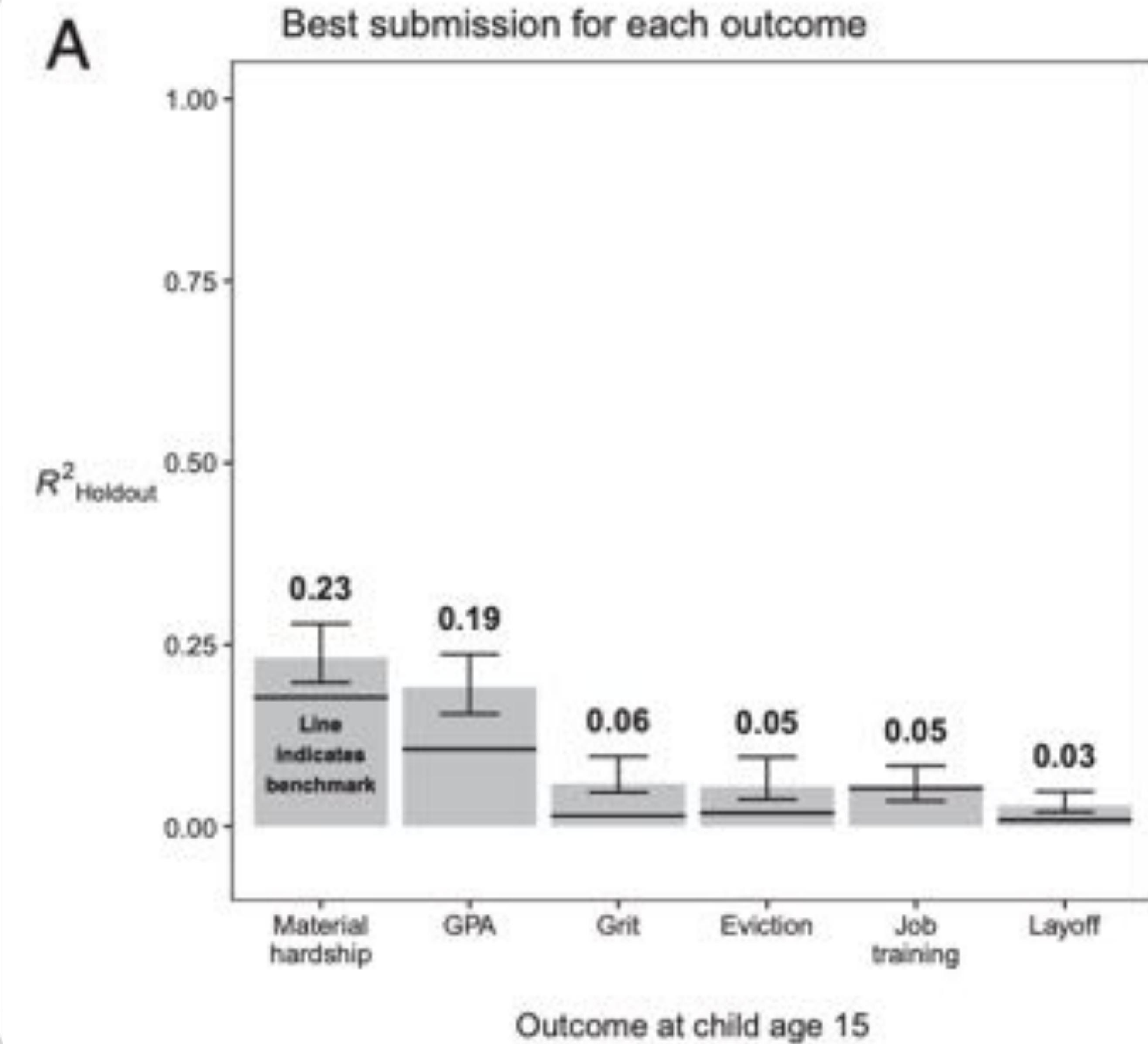
computing intensive

Predictability Crisis?

Measuring the predictability of life outcomes with a scientific mass collaboration

Matthew J. Salganik^{1,2}, Ian Lundberg³, Alexander T. Kindel⁴, Caitlin E. Ahearn⁵, Khaled Al-Ghoneim⁶, Abdullah Almaatouq^{4,7}, Drew M. Altshuler⁸, Jennie E. Brand^{9,10}, Nicole Bohme Carnegie¹⁰, Ryan James Compton¹¹, Debanjan Datta¹, Thomas Davidson¹, Anna Filippova¹, Connor Gilroy¹¹, Brian J. Goode¹², Faman Jahani¹³, Ridhi Kashyap^{1,14}, Antje Kirchner¹, Stephen McKay¹, Allison C. Morgan¹⁰, Alex Pentland¹, Kivan Polimis¹, Louis Raes¹⁵, Daniel E. Rigobon¹, Claudia V. Roberts⁷, Diana M. Stanescu¹, Yoshihiko Suhara¹⁶, Adnan Usmani¹⁷, Erik H. Wang¹, Muna Adem¹⁸, Abdulla Alhajri¹⁹, Bedoor AlShebli²⁰, Redwane Amin²¹, Ryan B. Amos¹, Lisa R. Argyle¹⁰, Livia Baer-Bositis²², Moritz Büchi²³, Bo-Ryehn Chung¹, William Eggert¹, Gregory Faletto²⁴, ZhiLin Fan¹, Jeremy Freese²⁵, Tejomay Gadgil²⁶, Josh Gagne²⁷, Yue Gao²⁸, Andrew Halpern-Manners²⁸, Sonia P. Hashmi²⁹, Sonia Hausen²⁹, Guanhua He²⁹, Kimberly Higuera²⁹, Bernie Hogan²⁹, Ilana M. Horwitz²⁹, Lisa M. Hummel²⁹, Naman Jain¹, Kun Jin¹⁰, David Jurgens¹², Patrick Kaminski^{29,30}, Areg Karapetyan^{29,31}, E. H. Kim²⁹, Ben Leizman¹, Naija Liu¹, Malte Möser¹, Andrew E. Mack¹, Mayank Mahajan⁷, Noah Mandell²⁹, Helge Marahrens²⁹, Diana Mercado-Garcia²⁹, Viola Mocz²⁹, Katarina Mueller-Gantell²⁹, Ahmed Musse²⁹, Qiankun Niu²⁹, William Nowak²⁹, Hamidreza Omidvar²⁹, Andrew Orr¹, Karen Ouyang¹, Katy M. Pinto²⁹, Ethan Porter²⁹, Kristin E. Porter²⁹, Crystal Qian¹, Tamkinat Rau²⁹, Anahit Sargsyan²⁹, Thomas Schaffner¹, Landon Schnabel²⁹, Bryan Schonfeld¹, Ben Sender²⁹, Jonathan D. Tang¹, Emma Turkov²⁹, Austin van Loon²⁹, Onur Varol^{29,32}, Xiafei Wang¹, Zhi Wang^{29,33}, Julia Wang¹, Flora Wang¹, Samantha Weissman¹, Kirstie Whitaker^{29,34}, Maria K. Wolters^{29,35}, Wei Lee Woon²⁹, James Wu^{29,36}, Catherine Wu¹, Kengran Yang²⁹, Jingwen Yin¹, Bingyu Zhao²⁹, Chenyun Zhu¹, Jeanne Brooks-Gunn^{29,37}, Barbara E. Engelhardt²⁹, Moritz Hardt²⁹, Dean Knox¹, Karen Levy²⁹, Arvind Narayanan¹, Brandon M. Stewart¹, Duncan J. Watts^{29,38,39,40,41}, and Sara McLanahan¹.

data challenge:
predicting life outcomes
based on ~6000 variables
by 160 teams
both theory- & data-driven



data challenge



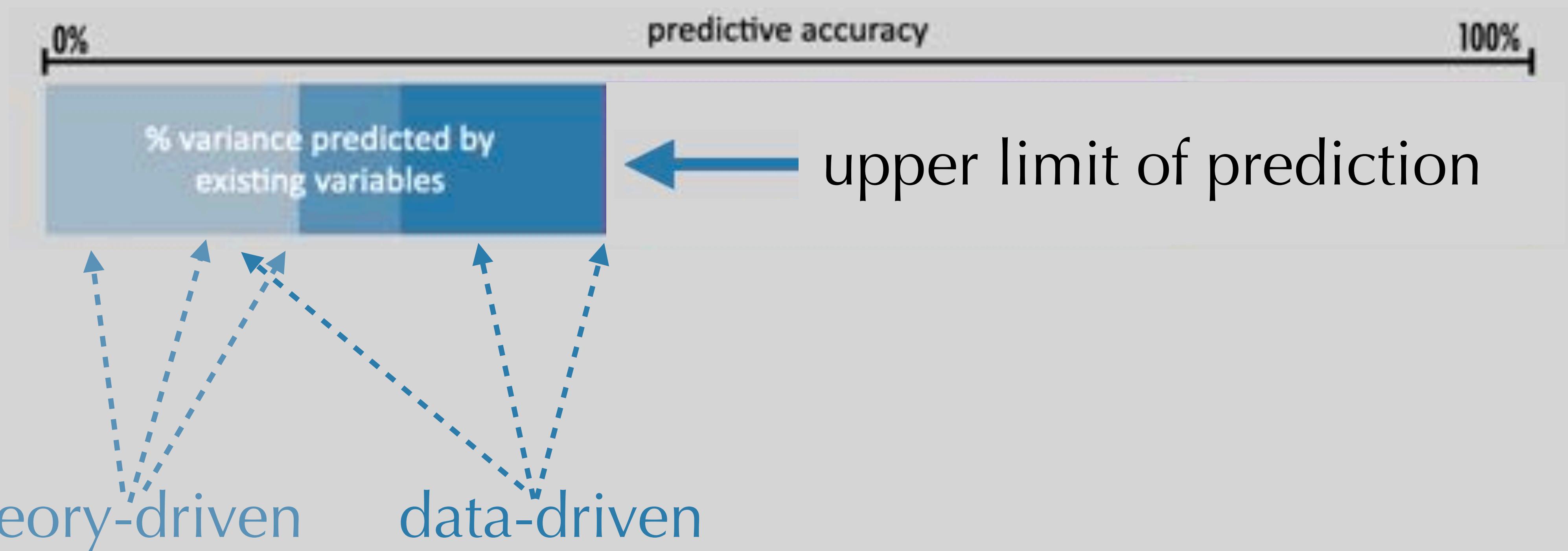
theory
driven

data
driven

theory- and data-driven teams
engage in common task
using common data
and common metric

Data Challenge

theory- and data-driven teams
engage in common task
using common data
and common metric

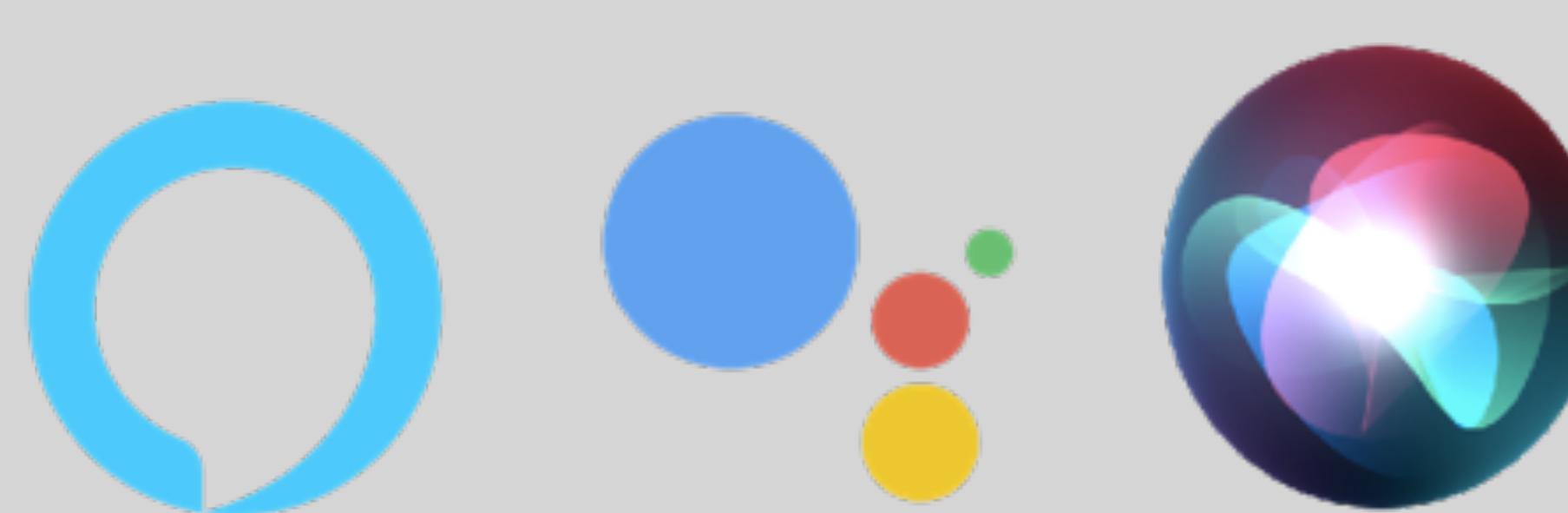


Prediction Benchmarks

“

Progress usually comes from many small improvements; a change of 1% can be a reason to break out the champagne

Liberman, 2012



🕒 Active Competitions

Hotness •



Google AI4Code - Understand Code in...

Predict the relationship between co...

Featured

Code Competition - 166 Teams

\$150,000

3 months to go



JPX Tokyo Stock Exchange Prediction

Explore the Tokyo market with your ...

Featured

Code Competition - 983 Teams

\$63,000

2 months to go



U.S. Patent Phrase to Phrase Matching

Help identify similar phrases in U.S. ...

Featured

Code Competition - 1258 Teams

\$25,000

8 months to go



Foursquare - Location Matching

Match point of interest data across ...

Featured

Code Competition - 489 Teams

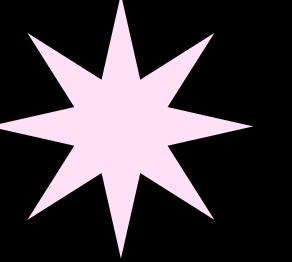
\$25,000

2 months to go

“

secret sauce of data science
Donoho, 2015

FERTILITY PREDICTION CHALLENGE



⌚ March-August 2024

University of Groningen,
Netherlands

0.54*

Is the current best [known to us] F1-score
of a classifier that predicts who is going
to have a child in the next three years

CAN YOU BEAT THIS SCORE?

Do you want to contribute to research on fertility behavior and the methodology of using prediction in social sciences?

Are you interested in working with unique registry-based datasets, including a social network for the entire Dutch population?

Are you looking for an engaging practical task for your machine learning course or workshop?

Or are you simply curious about the challenge and want to learn more about its design and prizes?



←
Sign up here to receive an update when the registration for the challenge opens and details are available

Contacts:

Gert Stulp g.stulp@rug.nl
Elizaveta Sivak e.sivak@rug.nl



university of
groningen



ODISSEI



Eyra



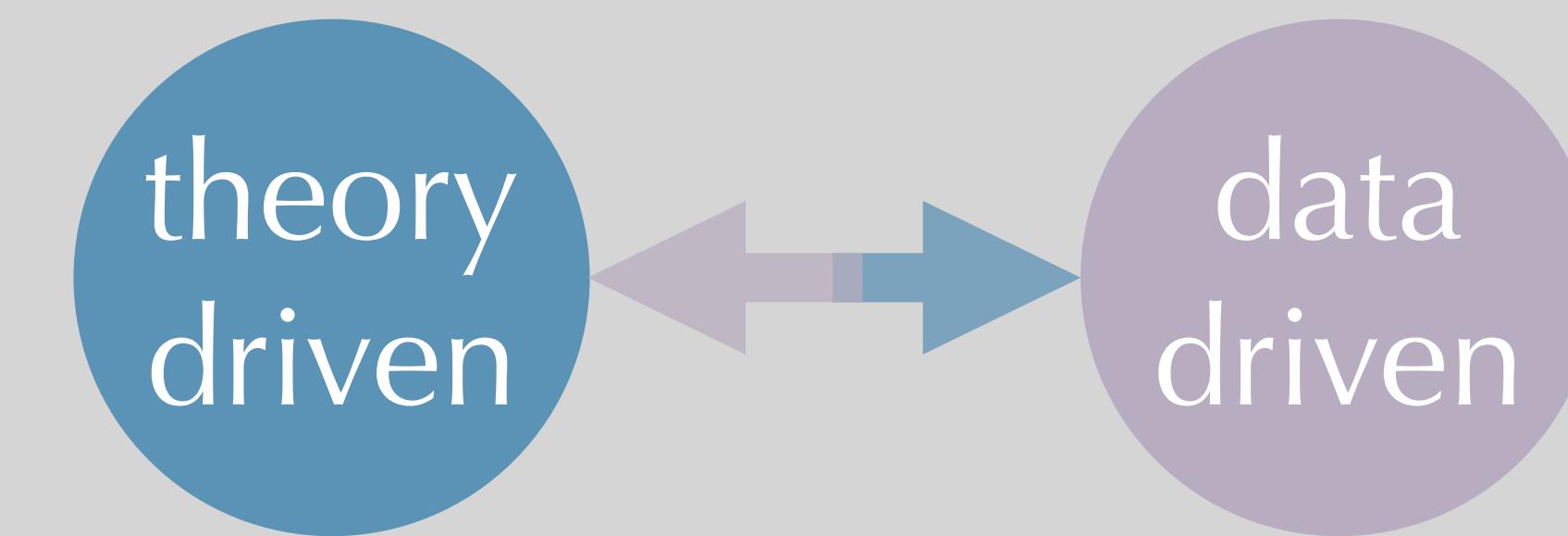
Take-Home Messages

a shift towards **prediction**
leads to a more reliable
and useful social science

out-of-sample predictive ability:



clear measure of
effect size



facilitates dialogue
theory- and data-
driven models



measure of distance
theory and practice



out-of-sample predictive ability
is a measure of how useful
our theory is in the real world

Article

The perils of policy by p-value: Predicting civil conflicts

Michael D Ward

Department of Political Science, Duke University

Brian D Greenhill

Department of Political Science, University of Washington

Kristin M Bakke

Department of Political Science, University College London



So Useful as a Good Theory? The Practicality Crisis in (Social) Psychological Theory

Elliot T. Berkman and Sylas M. Wilson

Department of Psychology and Center for Translational Neuroscience, University of Oregon



Frontiers in Psychological Science
3-12
© The Author(s) 2015.
Article reuse guidelines:
<http://journals.sagepub.com/page/journalsPermissions.nav>
www.psychologicalscience.org/permissions.nav
SAGE



out-of-sample predictive ability
is a measure of how useful
our theory is in the real world

 CrossMark
click for updates

Why significant variables aren't automatically good predictors

Adeline Lo^a, Herman Chernoff^{b,†}, Tian Zheng^c, and Shaw-Hwa Lo^{c,†}

^aDepartment of Political Science, University of California, San Diego, La Jolla, CA 92093; ^bDepartment of Statistics, Harvard University, Cambridge, MA 02138; and ^cDepartment of Statistics, Columbia University, New York, NY 10027

Contributed by Herman Chernoff, September 17, 2015 (sent for review December 15, 2014)

Thus far, genome-wide association studies (GWAS) have been disappointing in the inability of investigators to use the results of

From the scientist's point of view there are two basic problems, complicated by the large size of the data set. These are variable

PNAS



out-of-sample predictive ability
is a measure of how useful
our theory is in the real world

“

Social scientists studying the life course must find a way to reconcile a widespread belief that understanding has been generated by these data—as demonstrated by more than 750 published journal articles using the Fragile Families data with the fact that the very same data could not yield accurate predictions of these important outcomes.

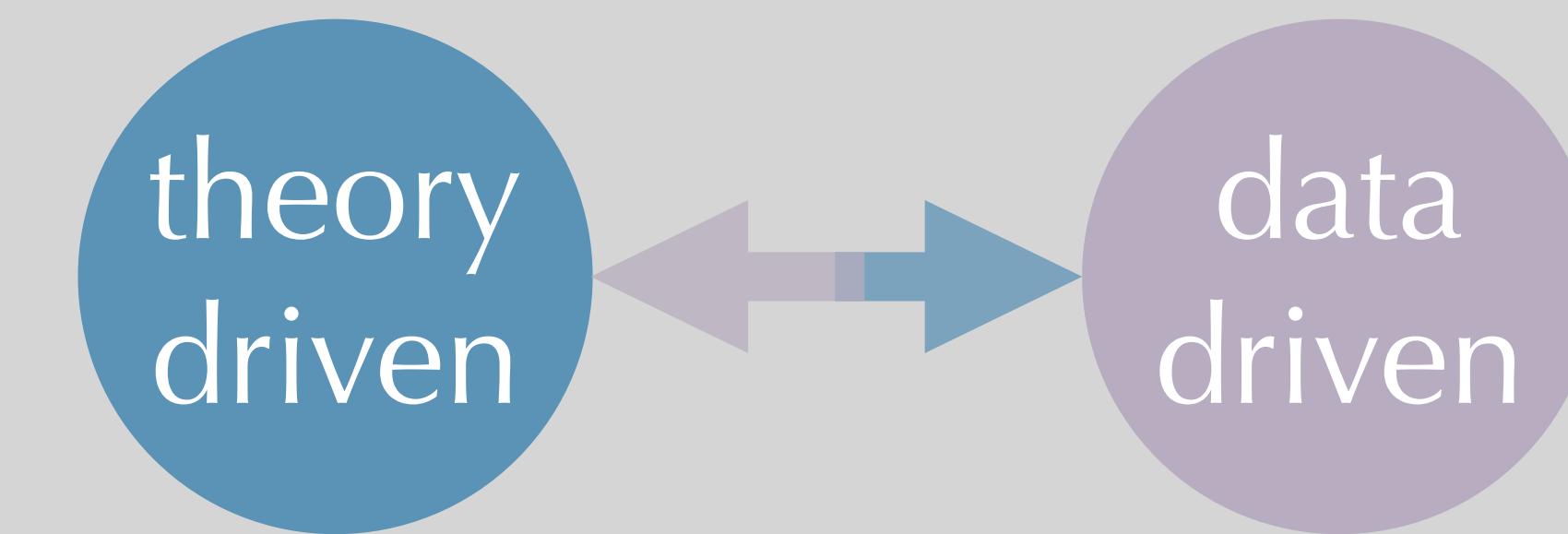
Take-Home Messages

a shift towards **prediction**
leads to a more reliable
and useful social science

out-of-sample predictive ability:



clear measure of
effect size



facilitates dialogue
theory- and data-
driven models



measure of distance
theory and practice

No Panacea

Patterns

CellPress
OPEN ACCESS

Article

Leakage and the reproducibility crisis in machine-learning-based science

Sayash Kapoor^{1,2,*} and Arvind Narayanan¹

¹Department of Computer Science and Center for Information Technology Policy, Princeton University, Princeton, NJ 08540, USA

²Lead contact

*Correspondence: sayashk@princeton.edu
<https://doi.org/10.1016/j.patter.2023.100804>

THE BIGGER PICTURE Machine learning (ML) is widely used across dozens of scientific fields. However, a common issue called “data leakage” can lead to errors in data analysis. We surveyed a variety of research that uses ML and found that data leakage affects at least 294 studies across 17 fields, leading to overoptimistic findings. We classified these errors into eight different types. We propose a solution: model info sheets that can be used to identify and prevent each of these eight types of leakage. We also tested the reproducibility of ML in a specific field: predicting civil wars, where complex ML models were thought to outperform traditional statistical models. Interestingly, when we corrected for data leakage, the supposed superiority of ML models disappeared: they did not perform any better than older methods. Our work serves as a cautionary note against taking results in ML-based science at face value.



Development/Pre-production: Data science output has been rolled out/validated across multiple domains/problems

But Much Needed

PNAS RESEARCH ARTICLE PSYCHOLOGICAL AND COGNITIVE SCIENCES OPEN ACCESS Check for updates

An illusion of predictability in scientific results: Even experts confuse inferential uncertainty and outcome variability

Sam Zhang¹, Patrick R. Heck², Michelle N. Meyer³, Christopher F. Chabris⁴, Daniel G. Goldstein⁵, and Jake M. Hofman^{6,7}

Edited by Elke Weber, Princeton University, Princeton, NJ; received February 22, 2023; accepted June 26, 2023

Traditionally, scientists have placed more emphasis on communicating inferential uncertainty (i.e., the precision of statistical estimates) compared to outcome variability (i.e., the predictability of individual outcomes). Here, we show that this can lead to sizable misperceptions about the implications of scientific results. Specifically, we present three preregistered, randomized experiments where participants saw the same scientific findings visualized as showing only inferential uncertainty, only outcome variability, or both and answered questions about the size and importance of findings they were shown. Our results, composed of responses from medical professionals, professional data scientists, and tenure-track faculty, show that the prevalent form of visualizing only inferential uncertainty can lead to significant overestimates of treatment effects, even among highly trained experts. In contrast, we find that depicting both inferential uncertainty and outcome variability leads to more accurate perceptions of results while appearing to leave other subjective impressions of the results unchanged, on average.

statistics | uncertainty | science communication | visualization | experiments

The figure consists of two panels. The left panel is a dot plot with error bars showing aggression scores for 'Violent game' and 'Non-violent game'. The y-axis is labeled 'Aggressiveness score' and ranges from 6.6 to 6.9. The 'Violent game' dot is at approximately 6.82 with an error bar from 6.78 to 6.86. The 'Non-violent game' dot is at approximately 6.64 with an error bar from 6.61 to 6.67. The right panel is a density scatter plot showing aggression scores for 'Violent game' (purple) and 'Non-violent game' (green). The y-axis ranges from 5 to 8. Both distributions are centered around an aggression score of 7.

The Proposal

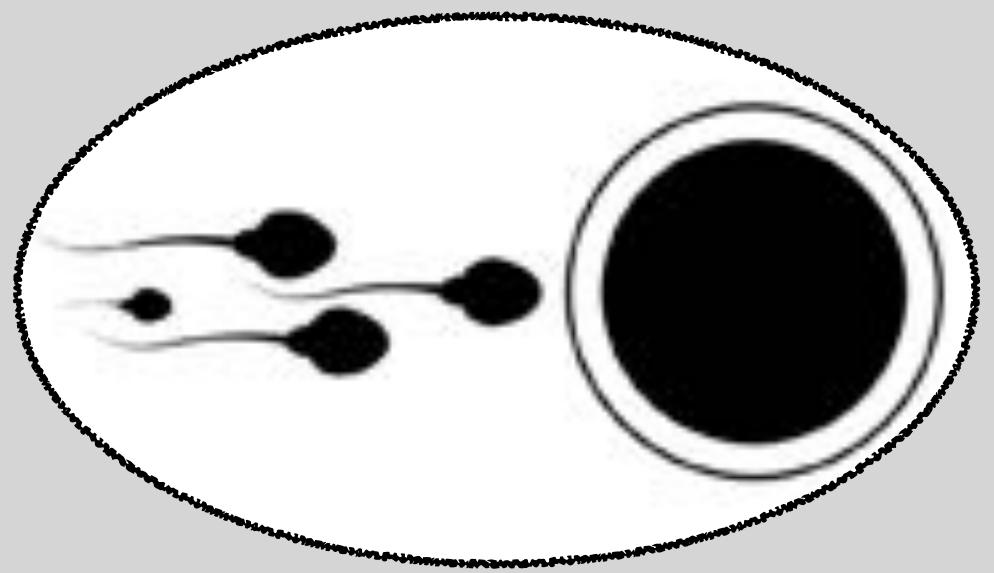
a shift towards **prediction**
leads to a more reliable
and useful social science

microsimulation can
advance traditional
statistical modelling

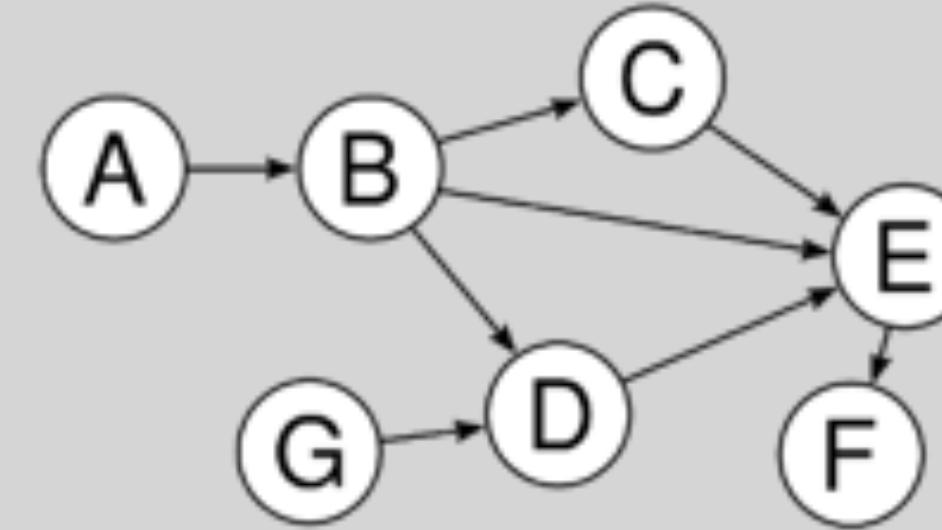
Take-Home Messages

microsimulation can
advance traditional
statistical modelling

microsimulation can:



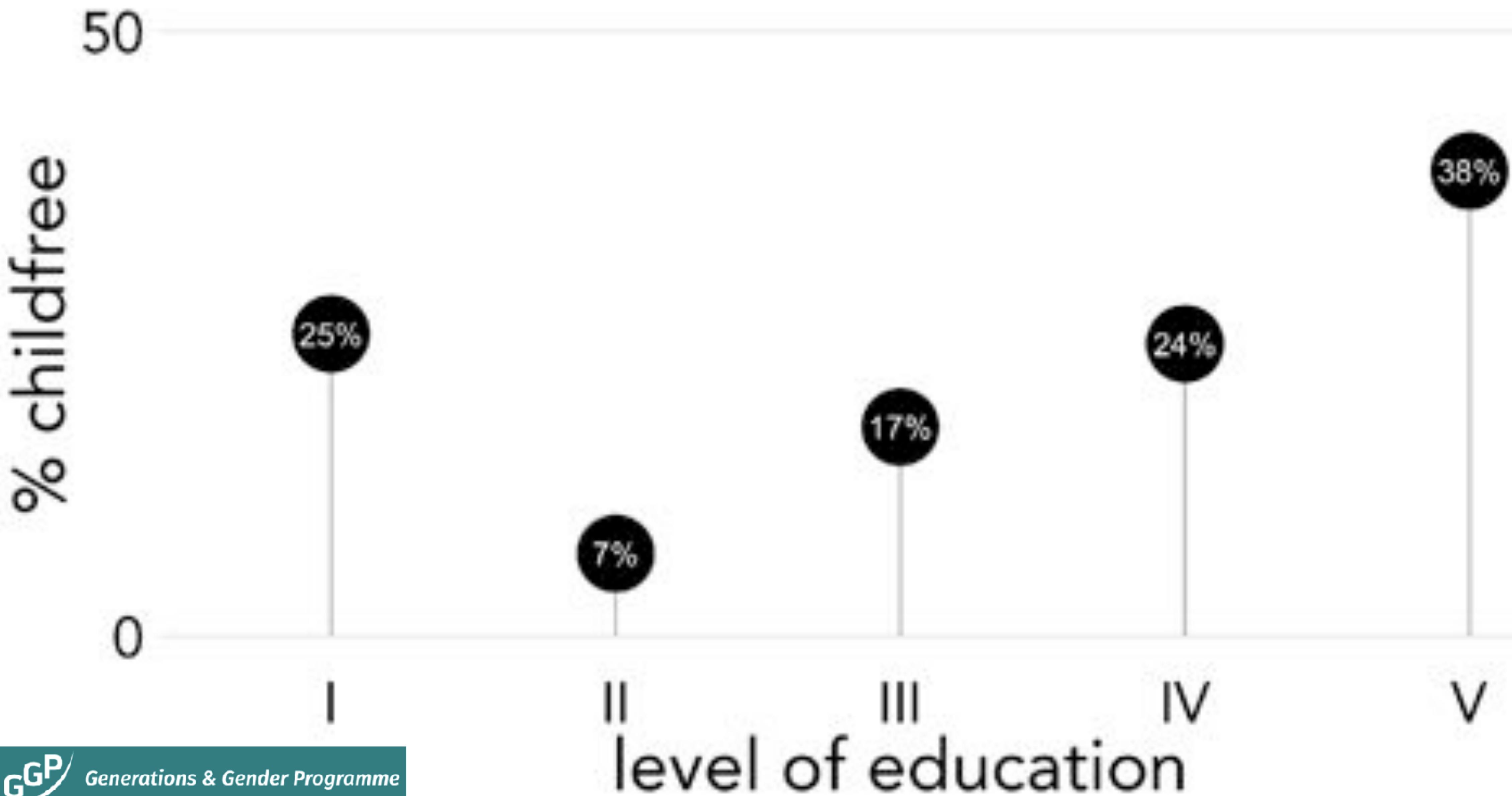
include
biological
information

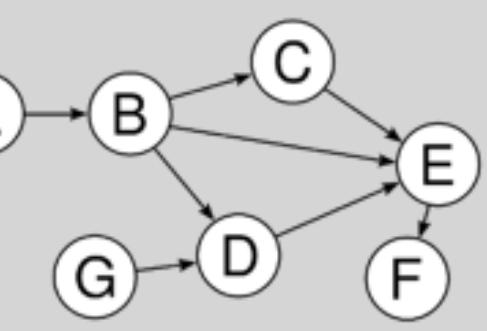
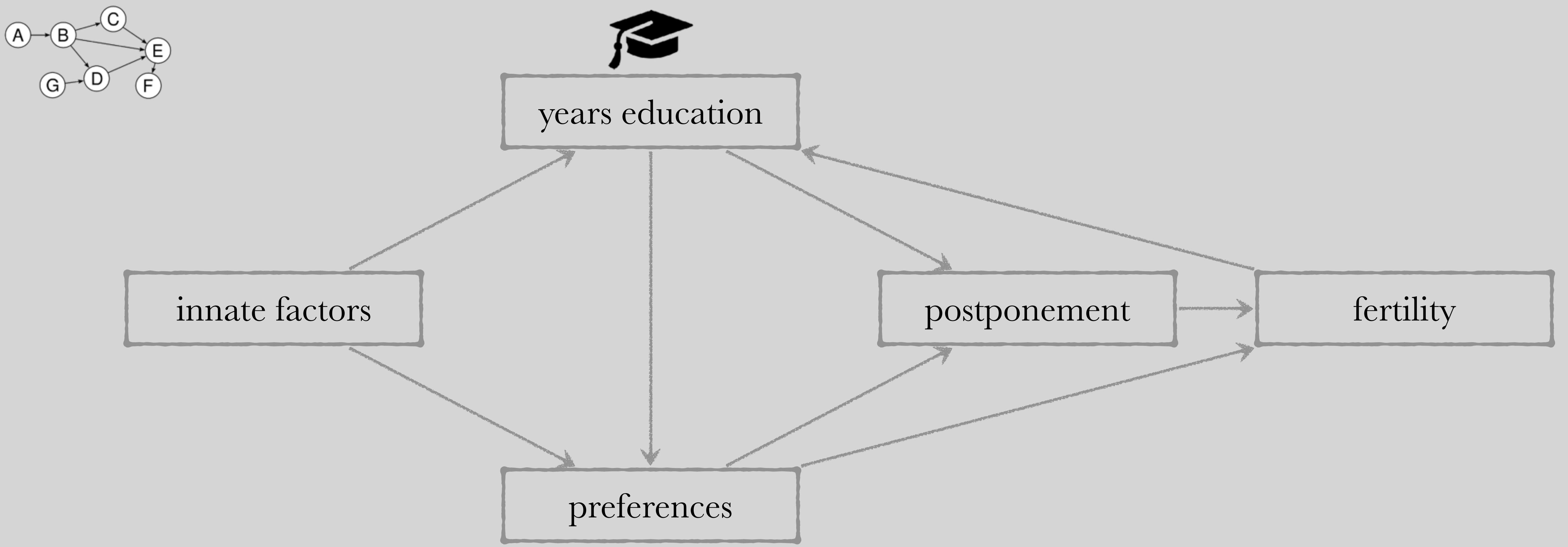


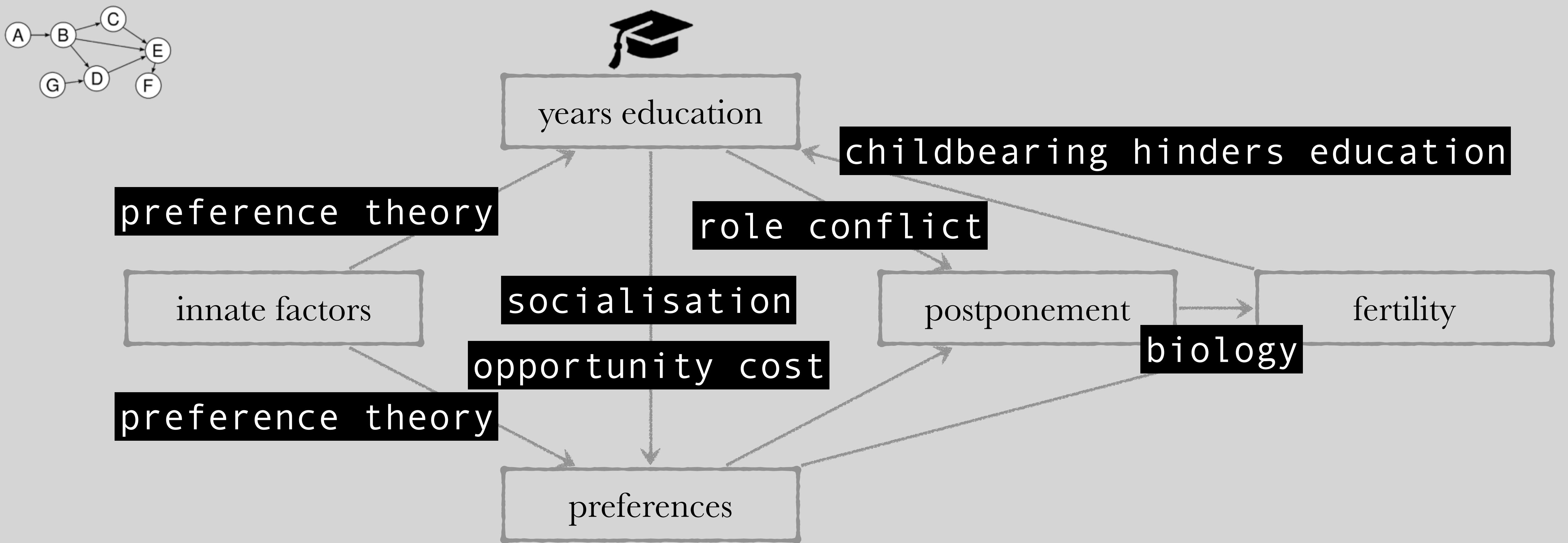
test (causal)
mechanisms

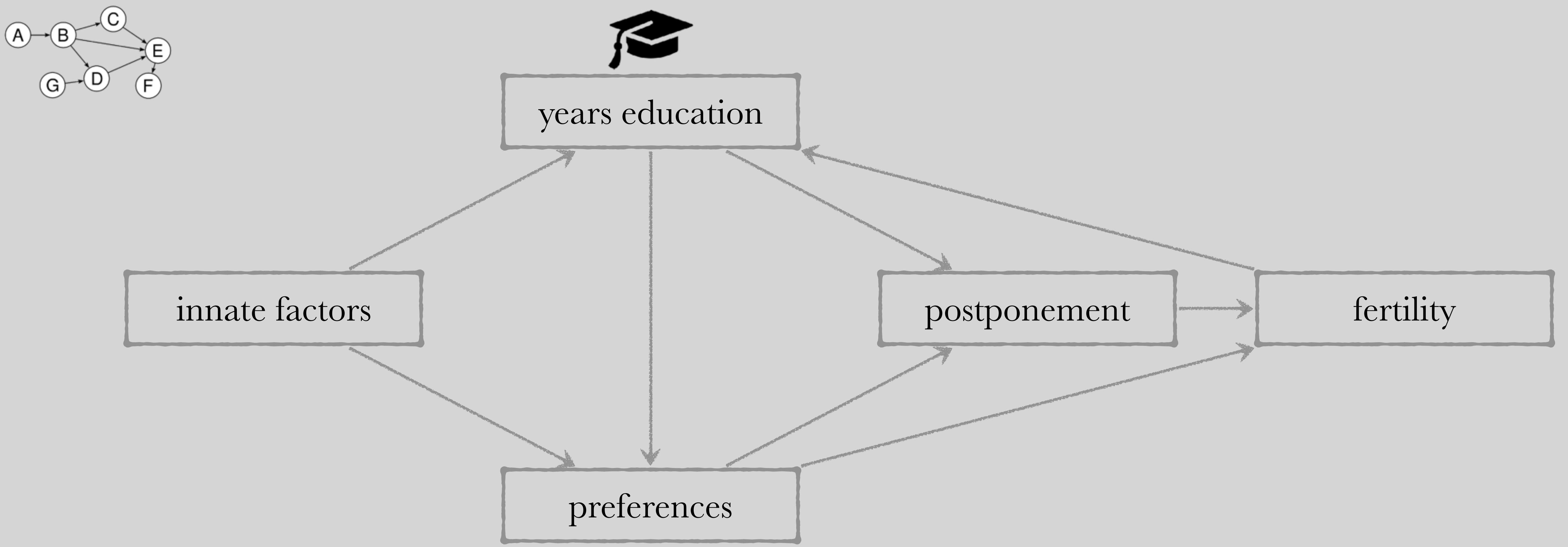


quantify
unpredictability



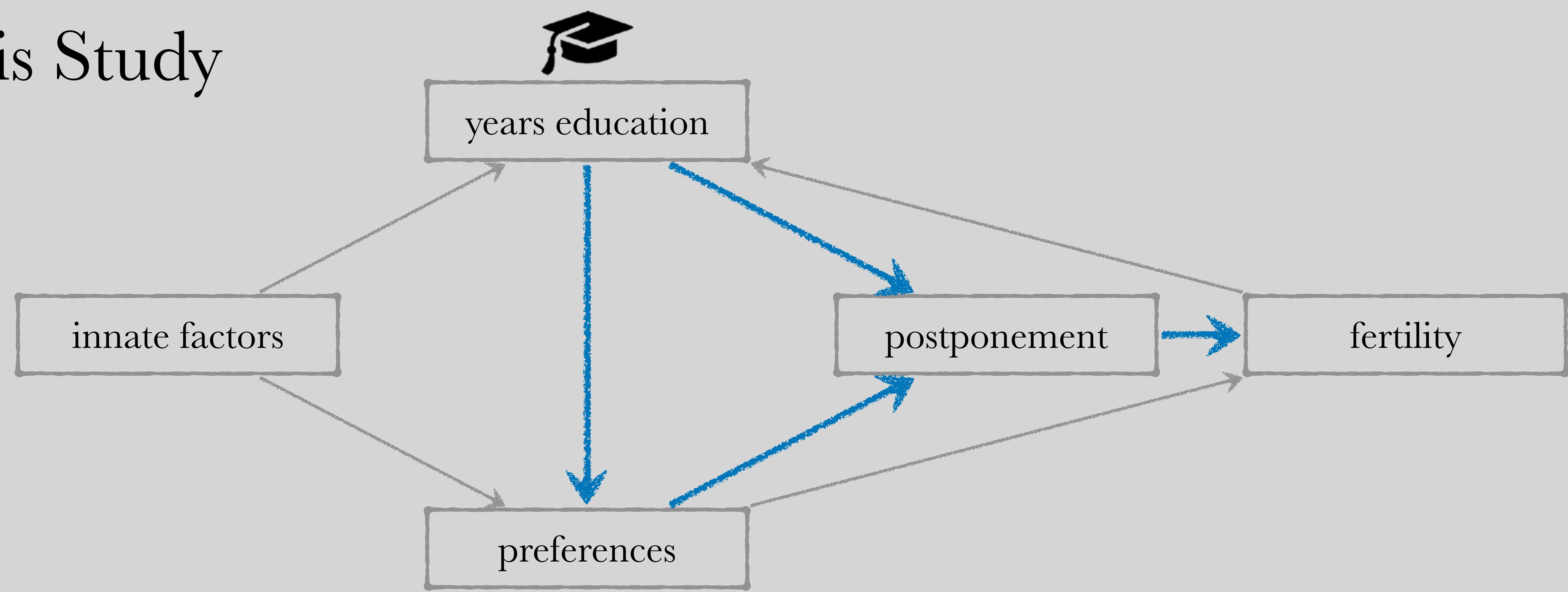




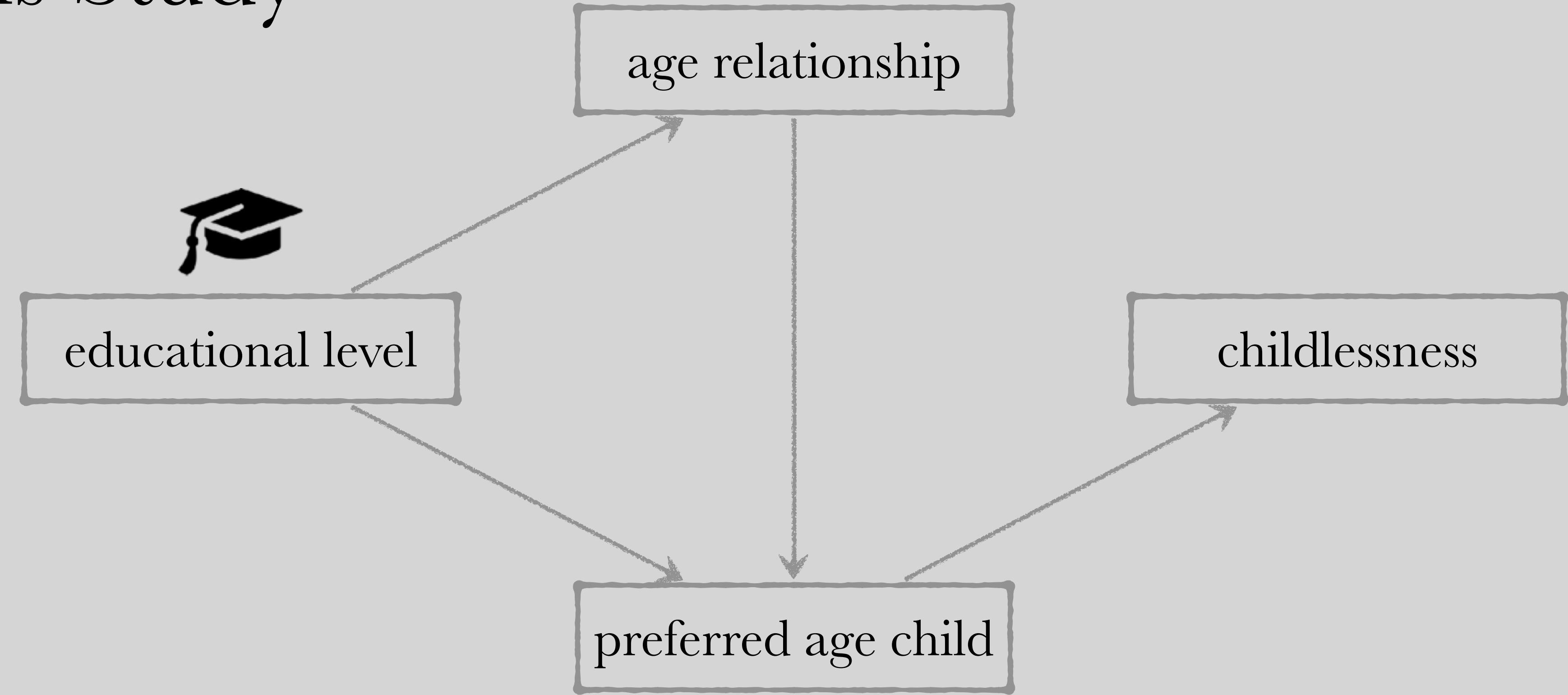


What Kind of Data
Would We need to
Address This Model?

This Study



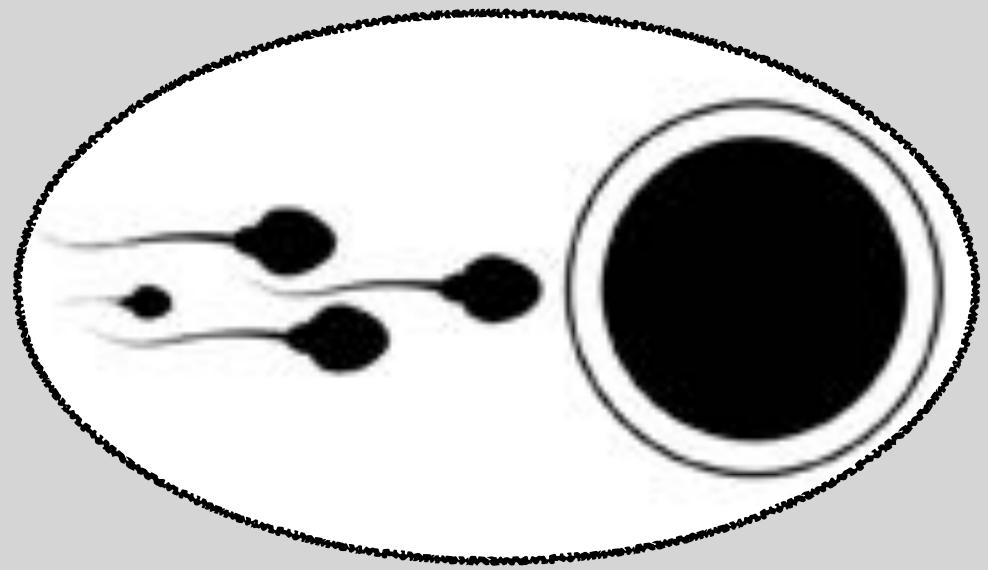
This Study



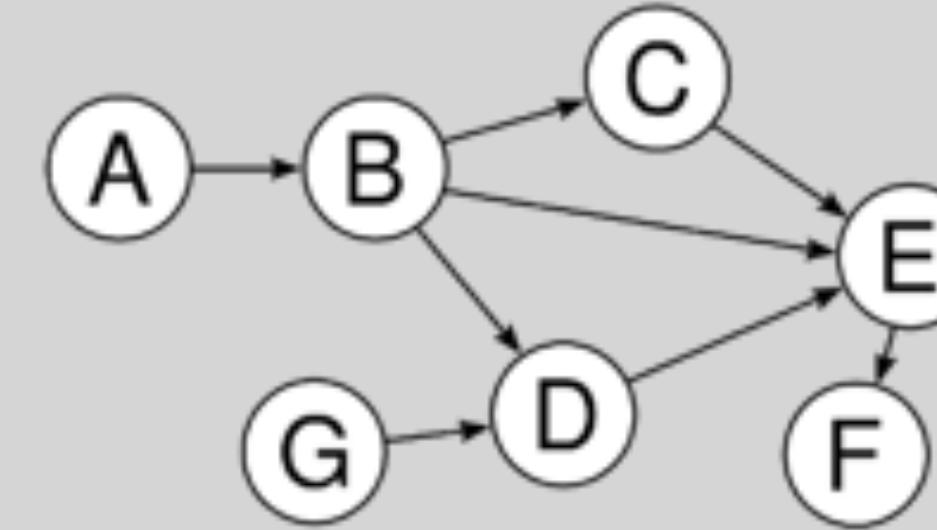
Take-Home Messages

microsimulation can
advance traditional
statistical modelling

microsimulation can:



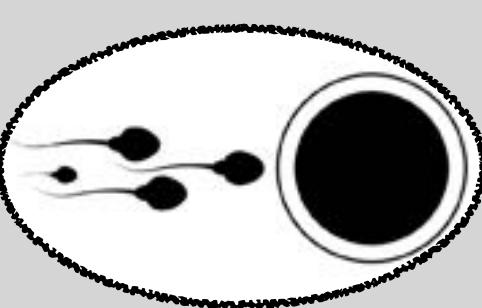
include
biological
information



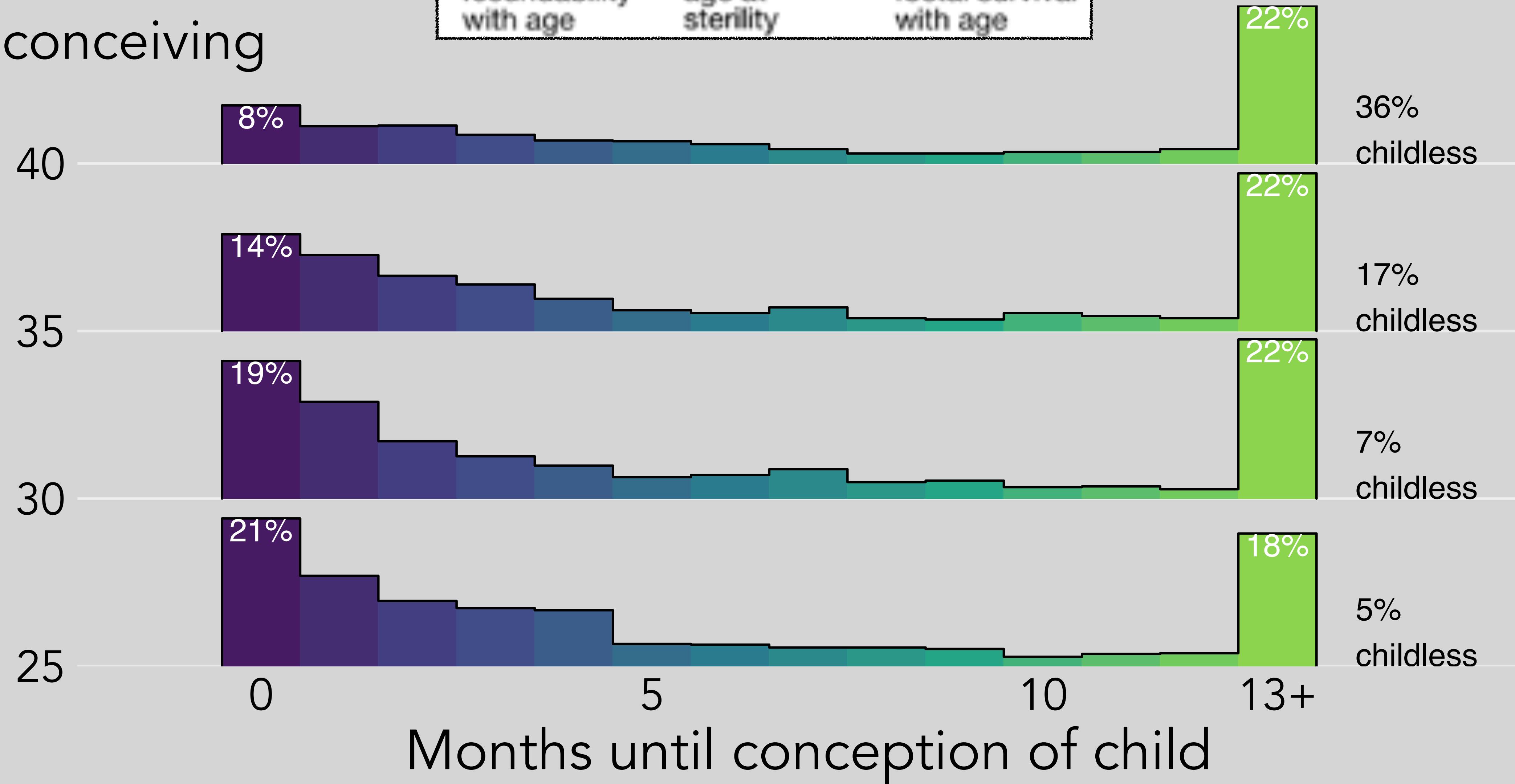
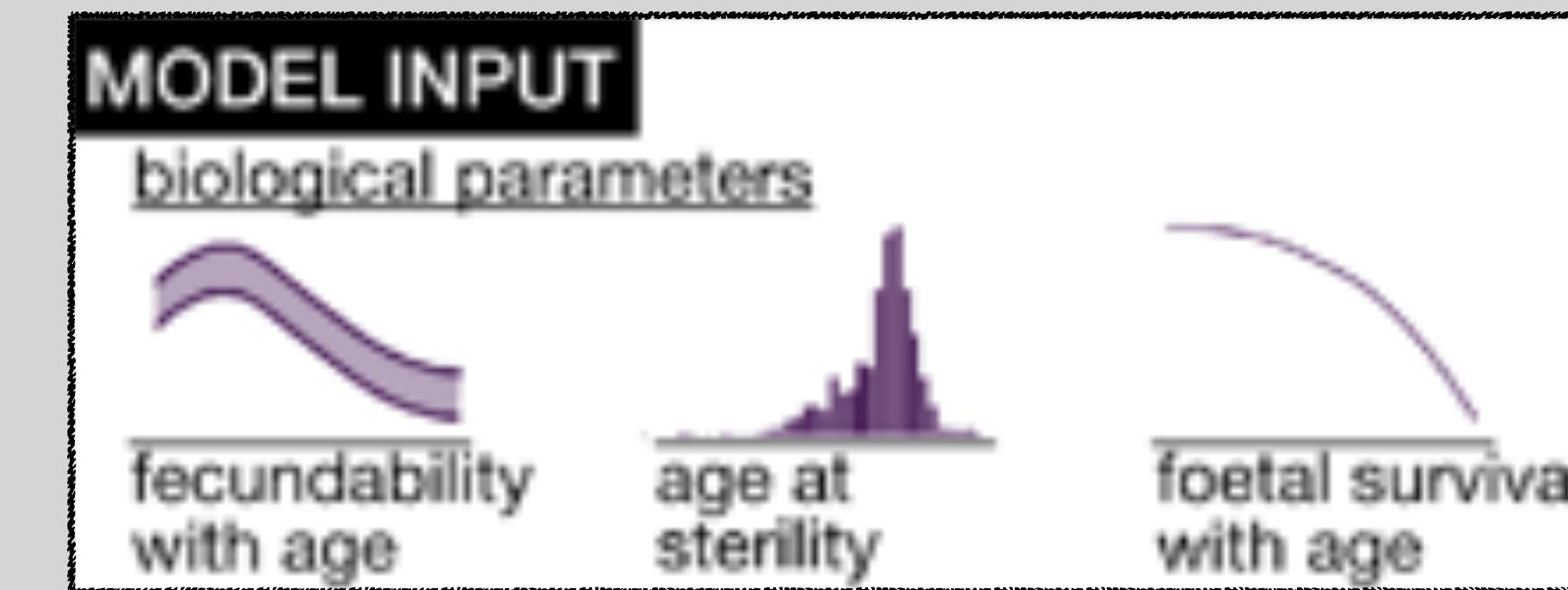
test (causal)
mechanisms

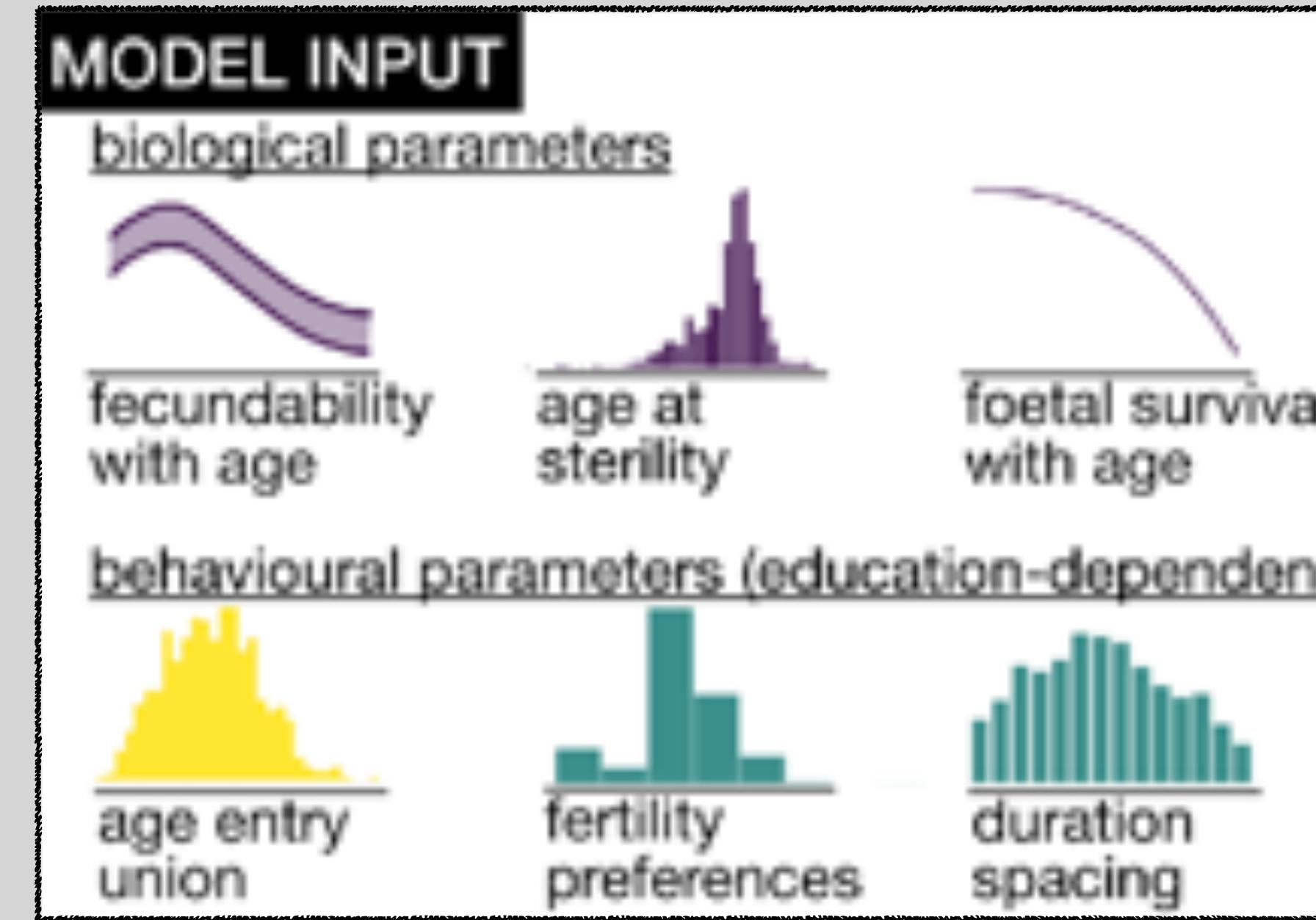
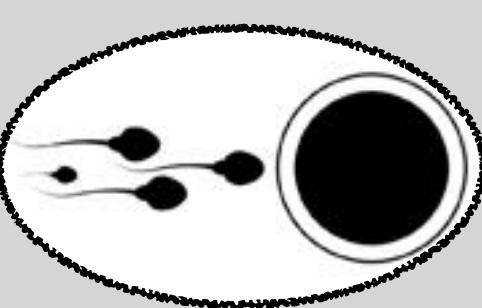


quantify
unpredictability



Age start conceiving





determines whether and when
people would like to conceive

determines whether and when
people conceive

MODEL INPUT

biological parameters



fecundability
with age



age at
sterility



foetal survival
with age

behavioural parameters (education-dependent)



age entry
union



fertility
preferences



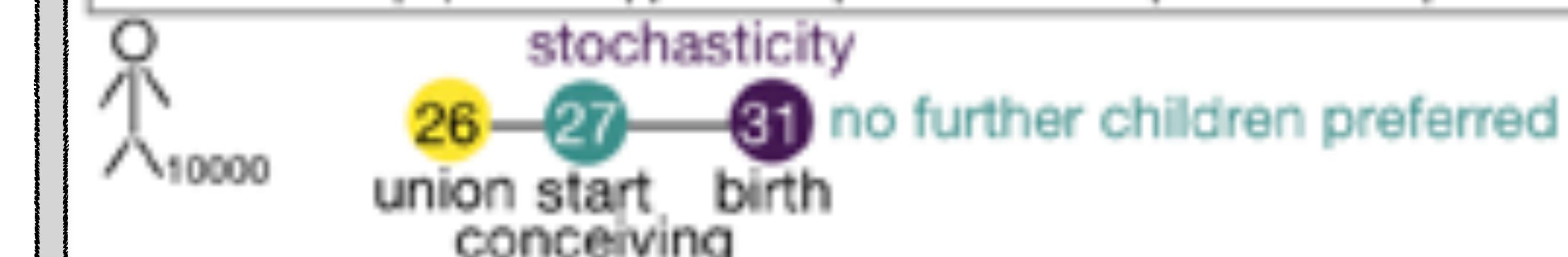
duration
spacing

MODEL RUN

Randomly determined traits individual 1
in union =22 | spac. =5 | pref. =2 | fecund. =0.3 | steril. =43 | edu. =high



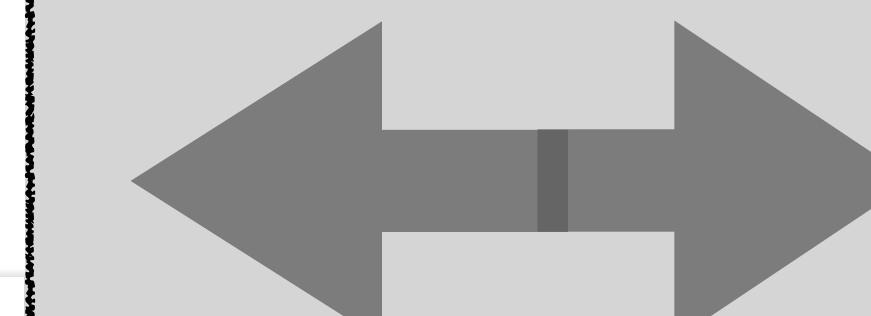
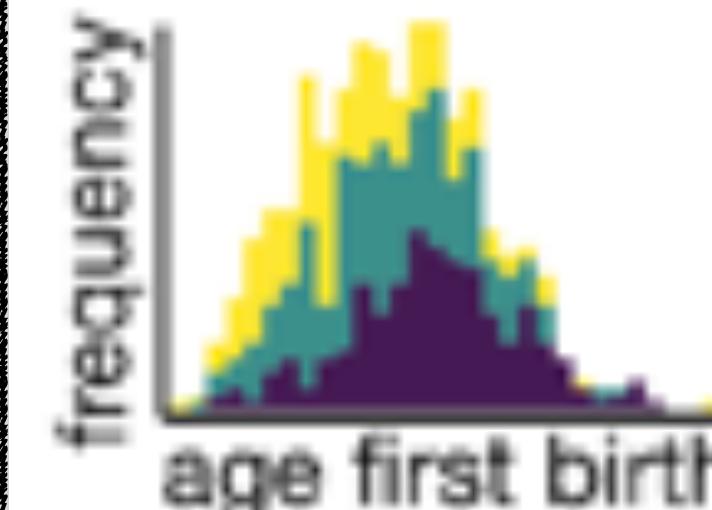
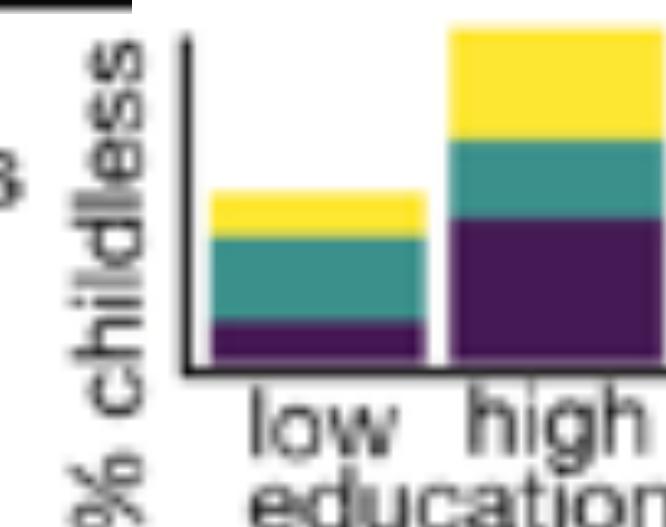
Randomly determined traits individual 10000
in union =26 | spac. =1 | pref. =1 | fecund. =0.1 | steril. =45 | edu. =low



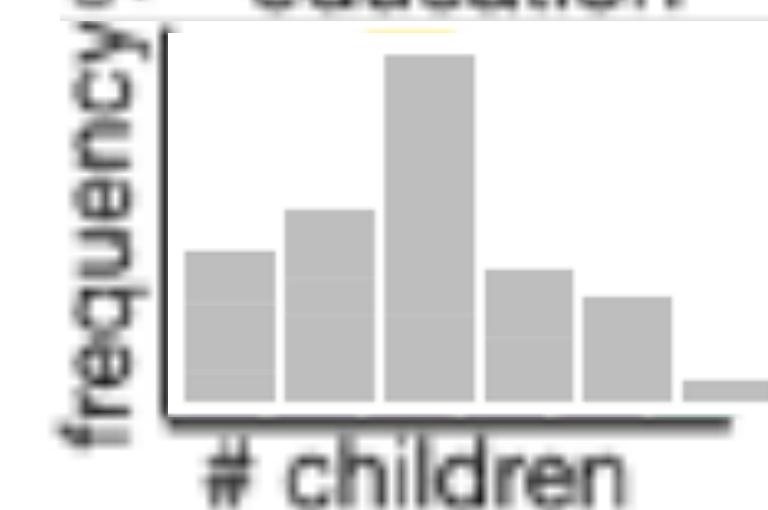
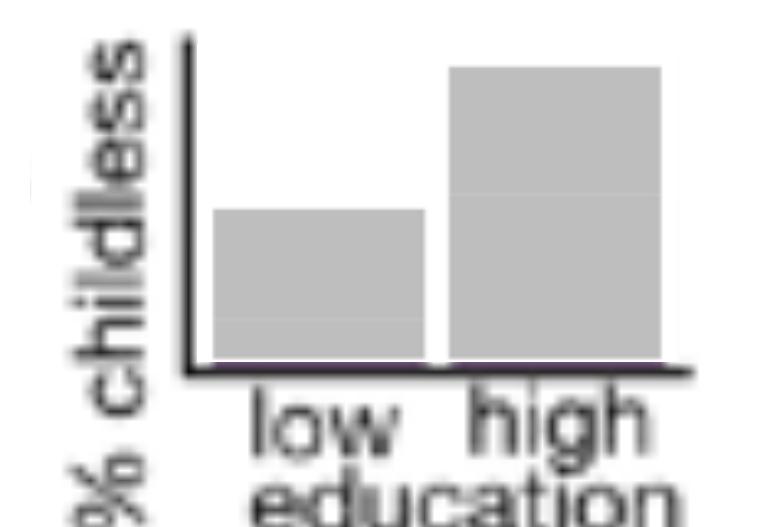
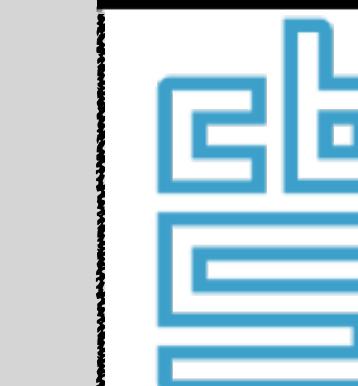
MODEL OUTPUT

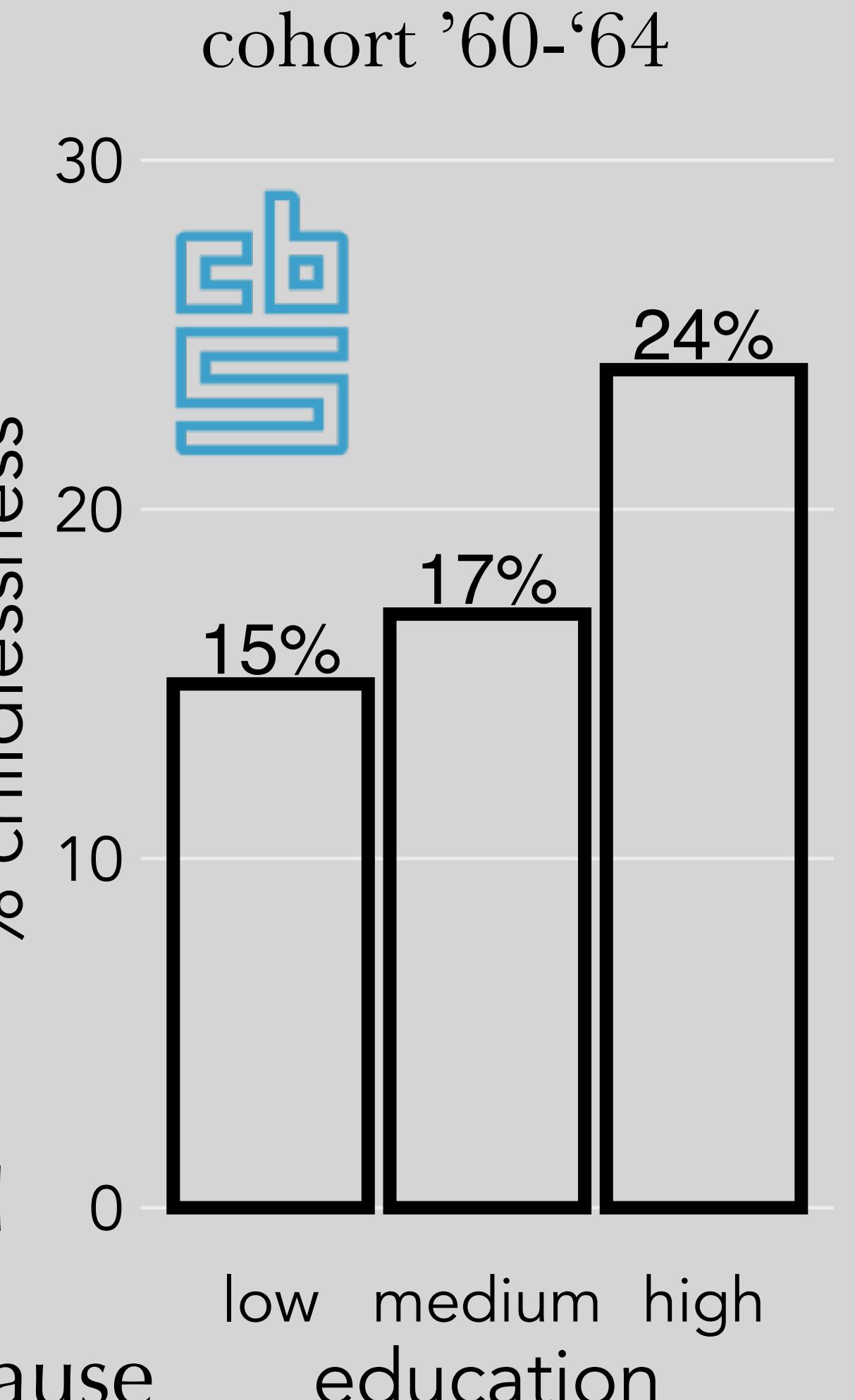
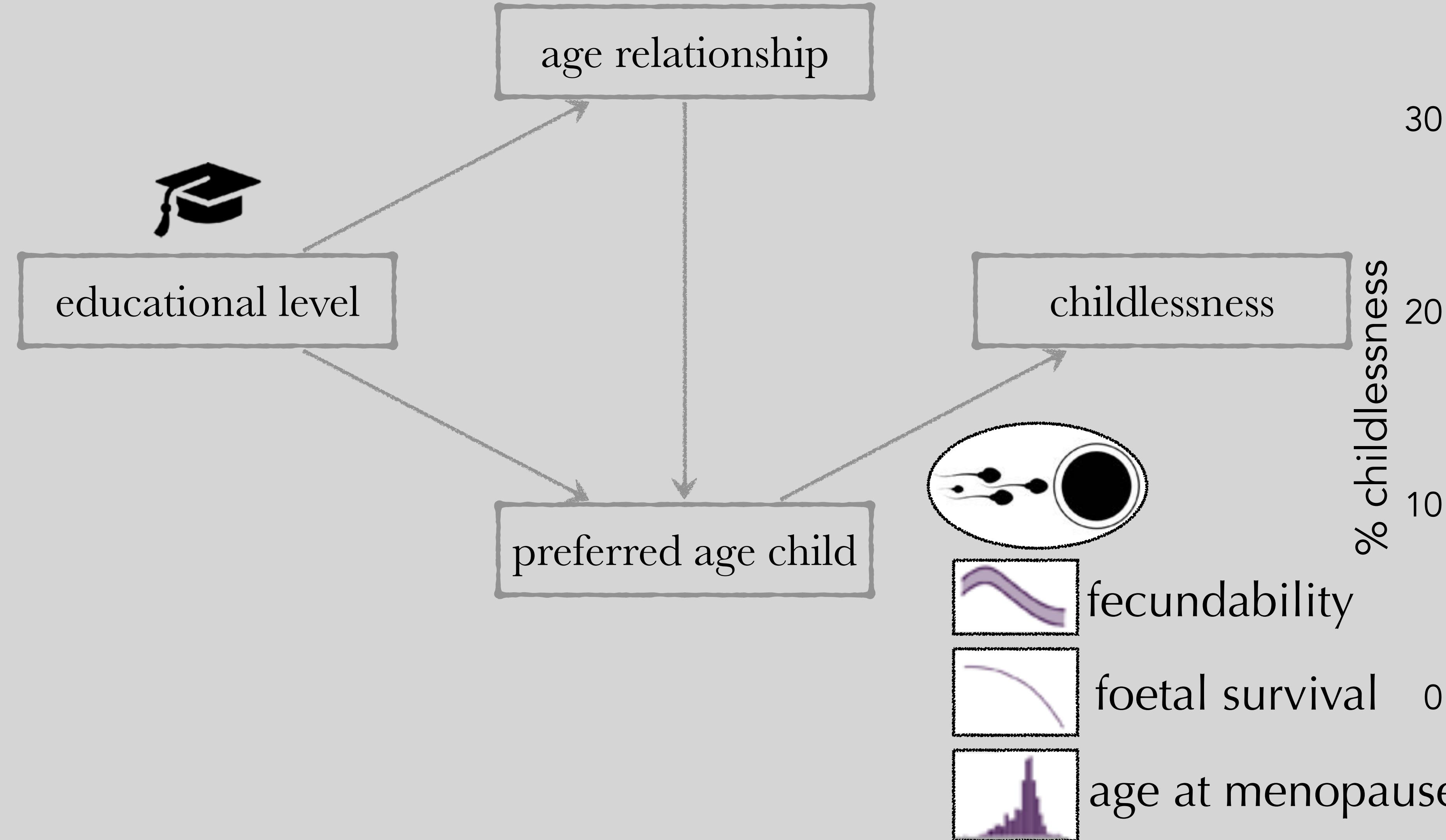
due to:

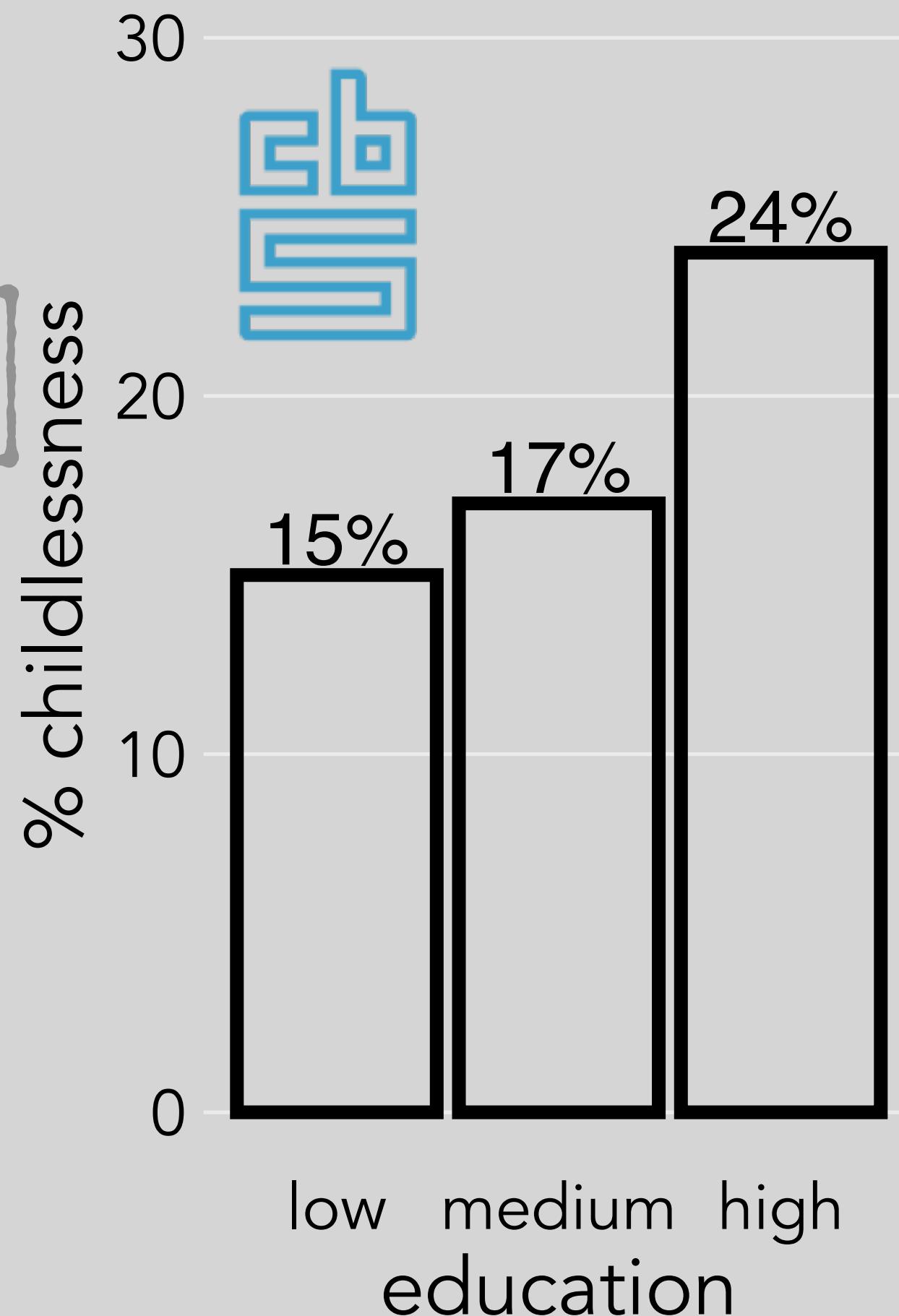
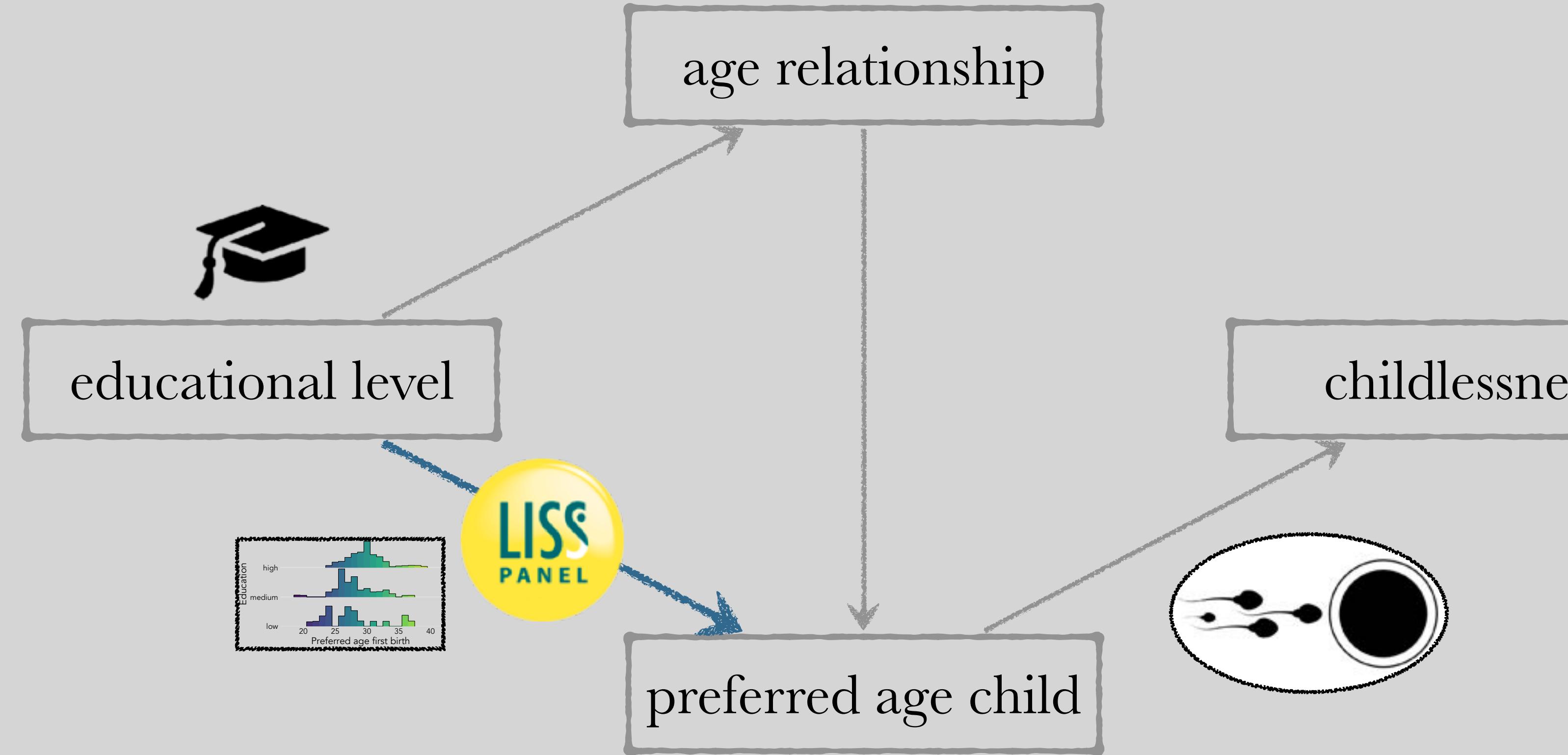
- partner status
- preferences
- stochasticity

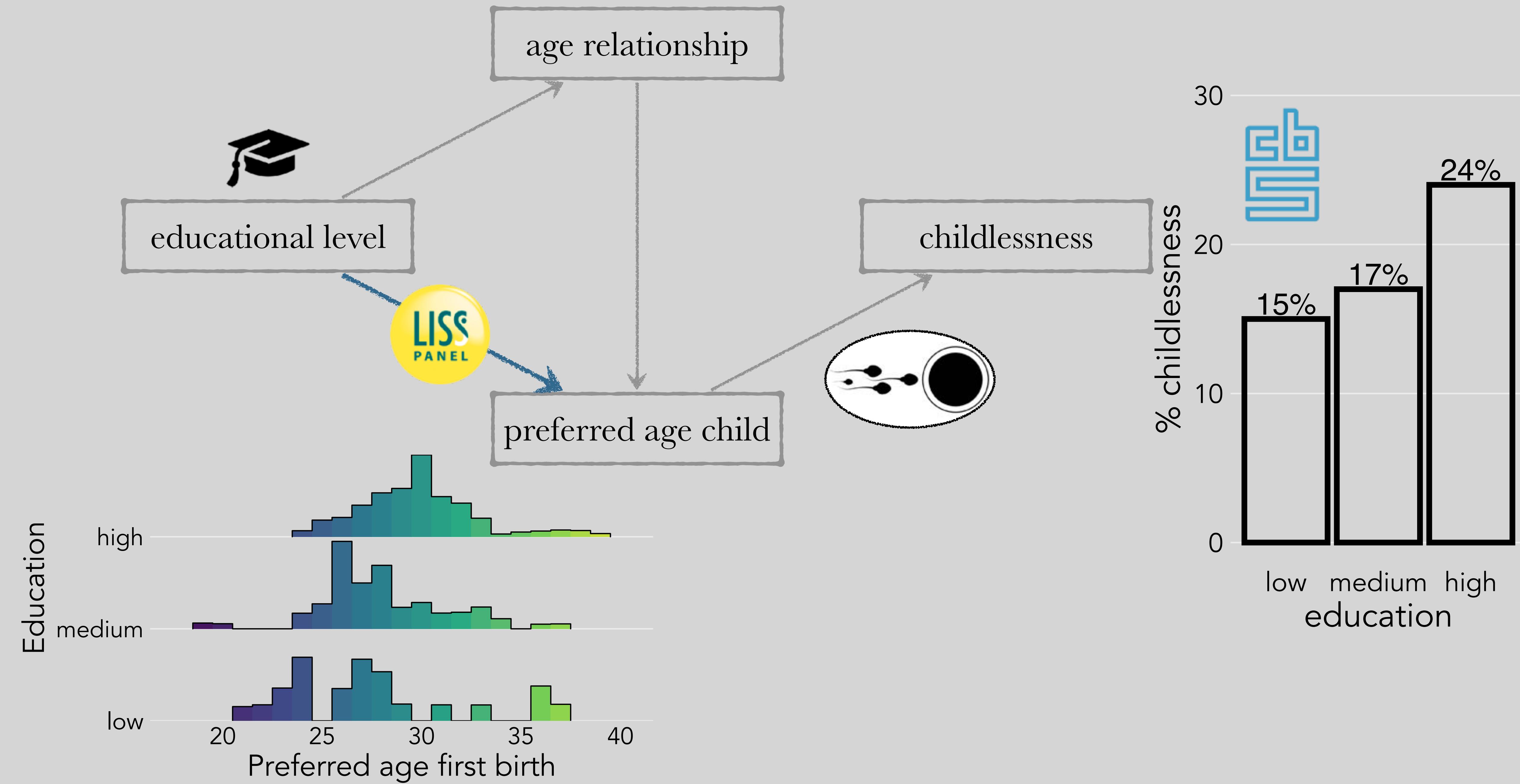


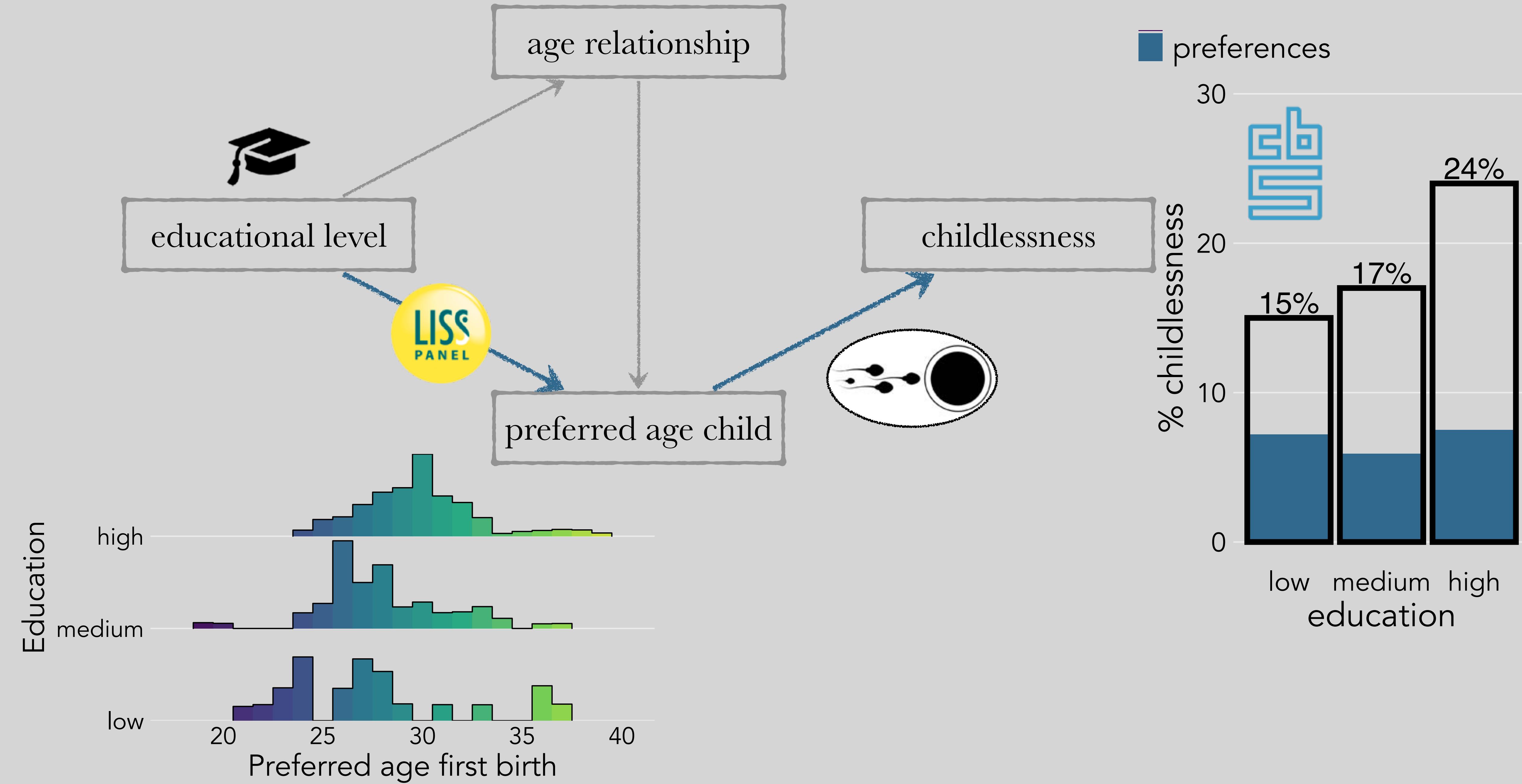
'TRUTH'

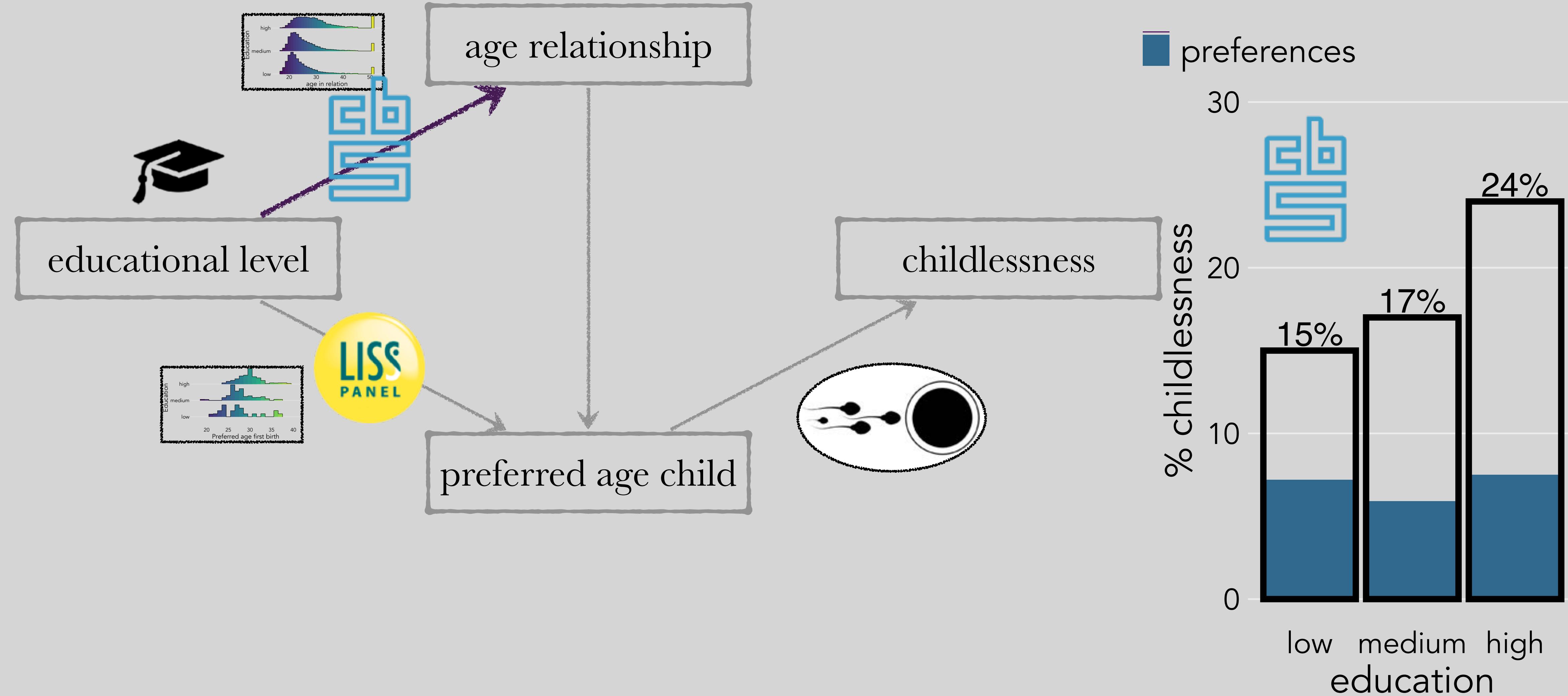


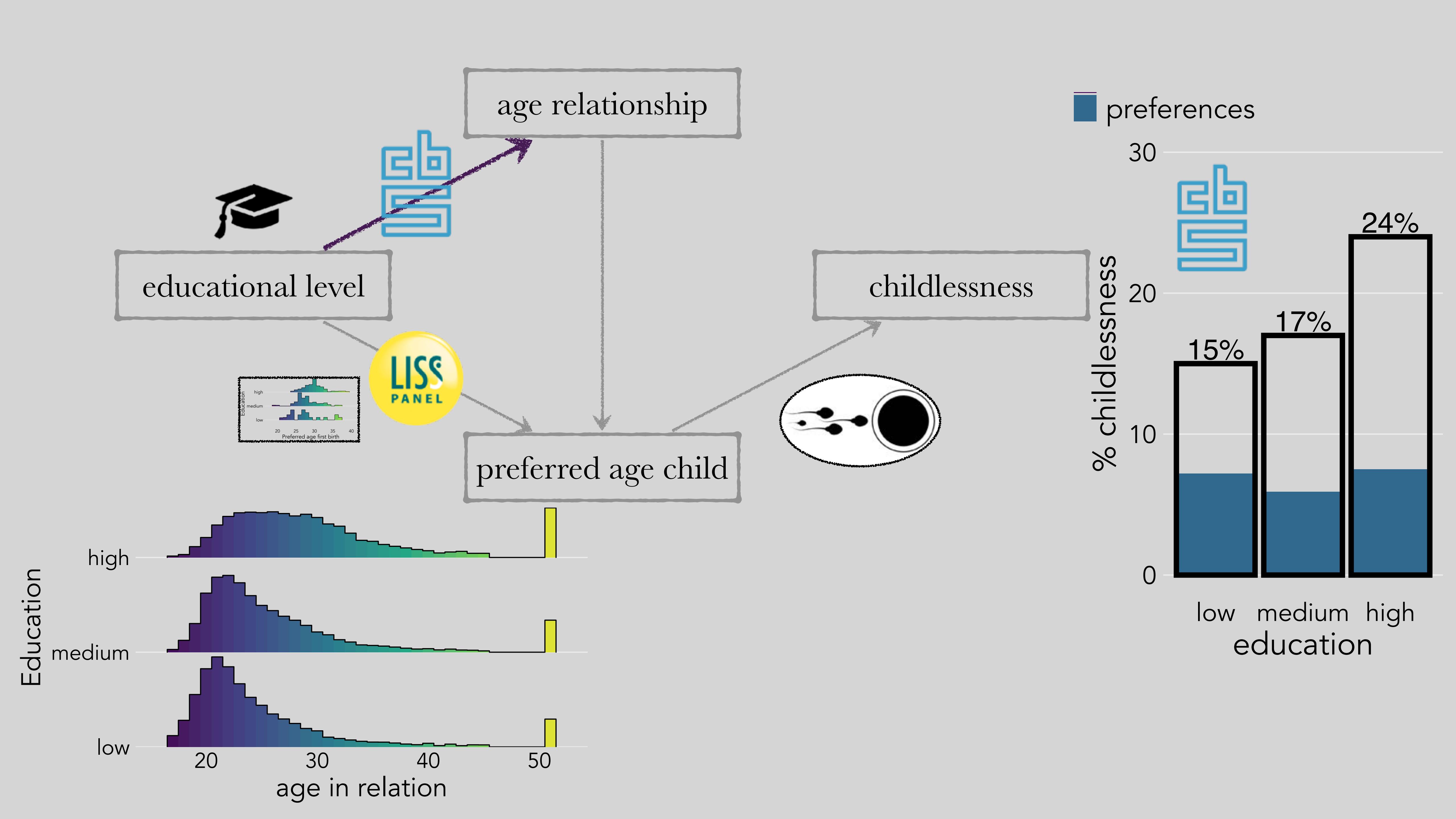


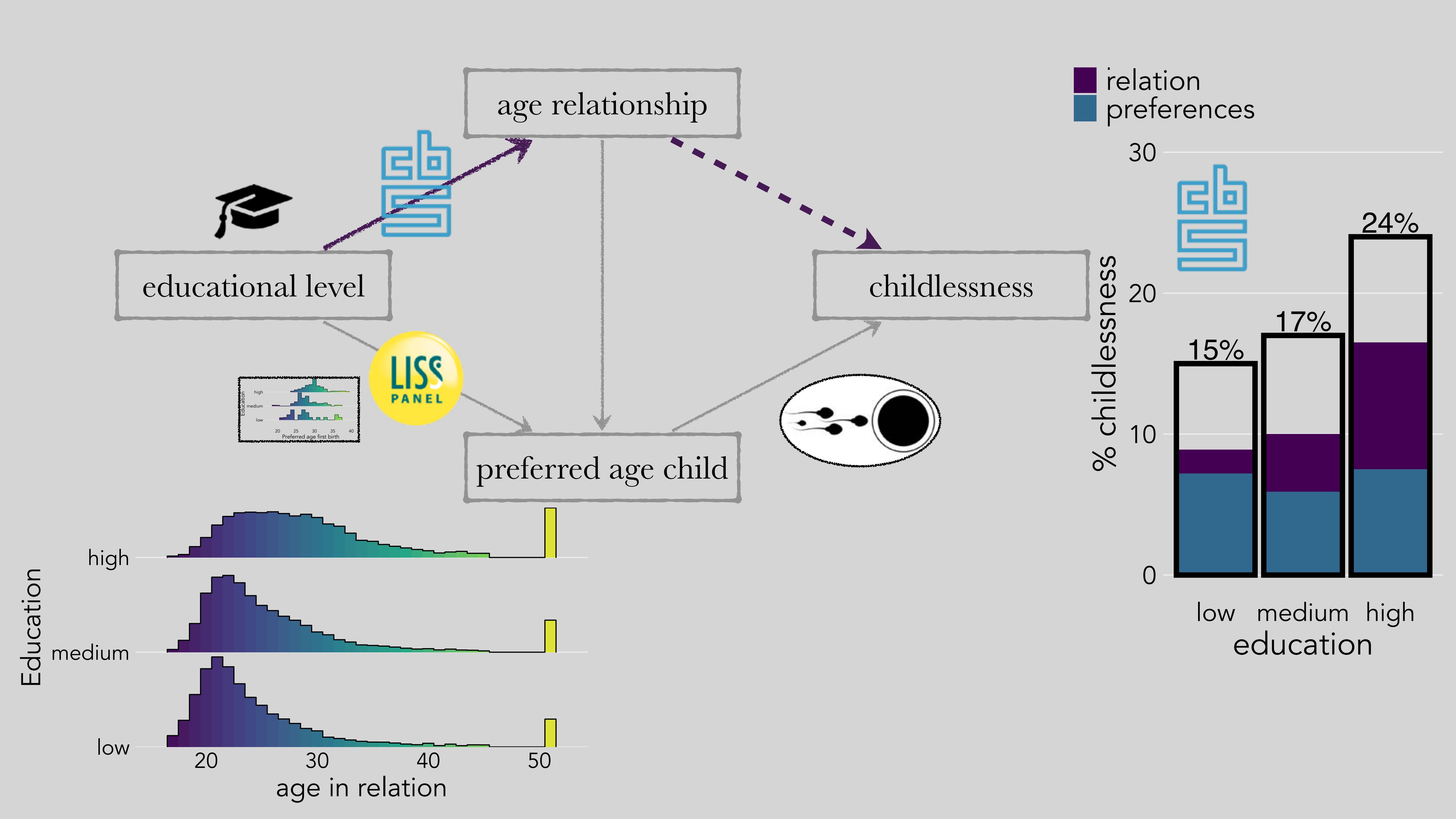




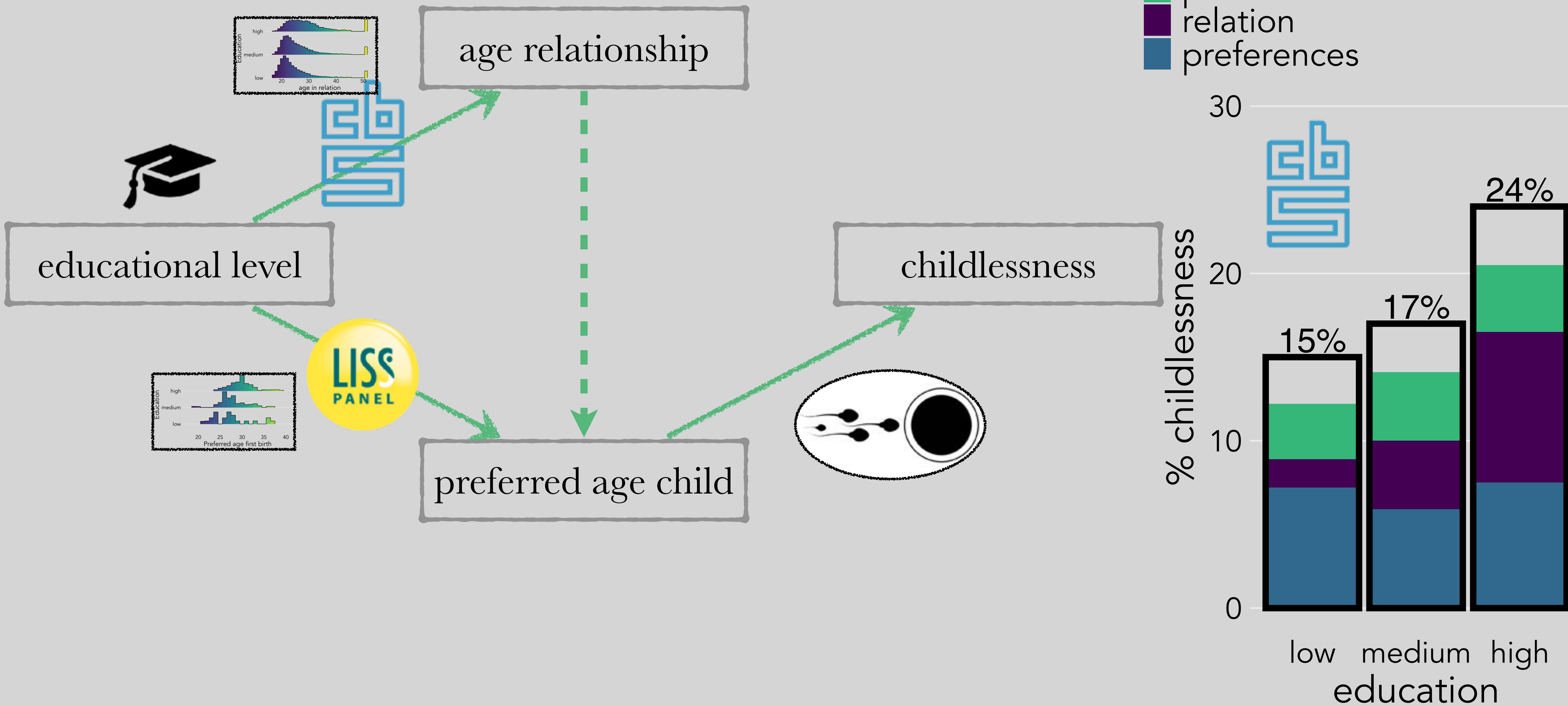


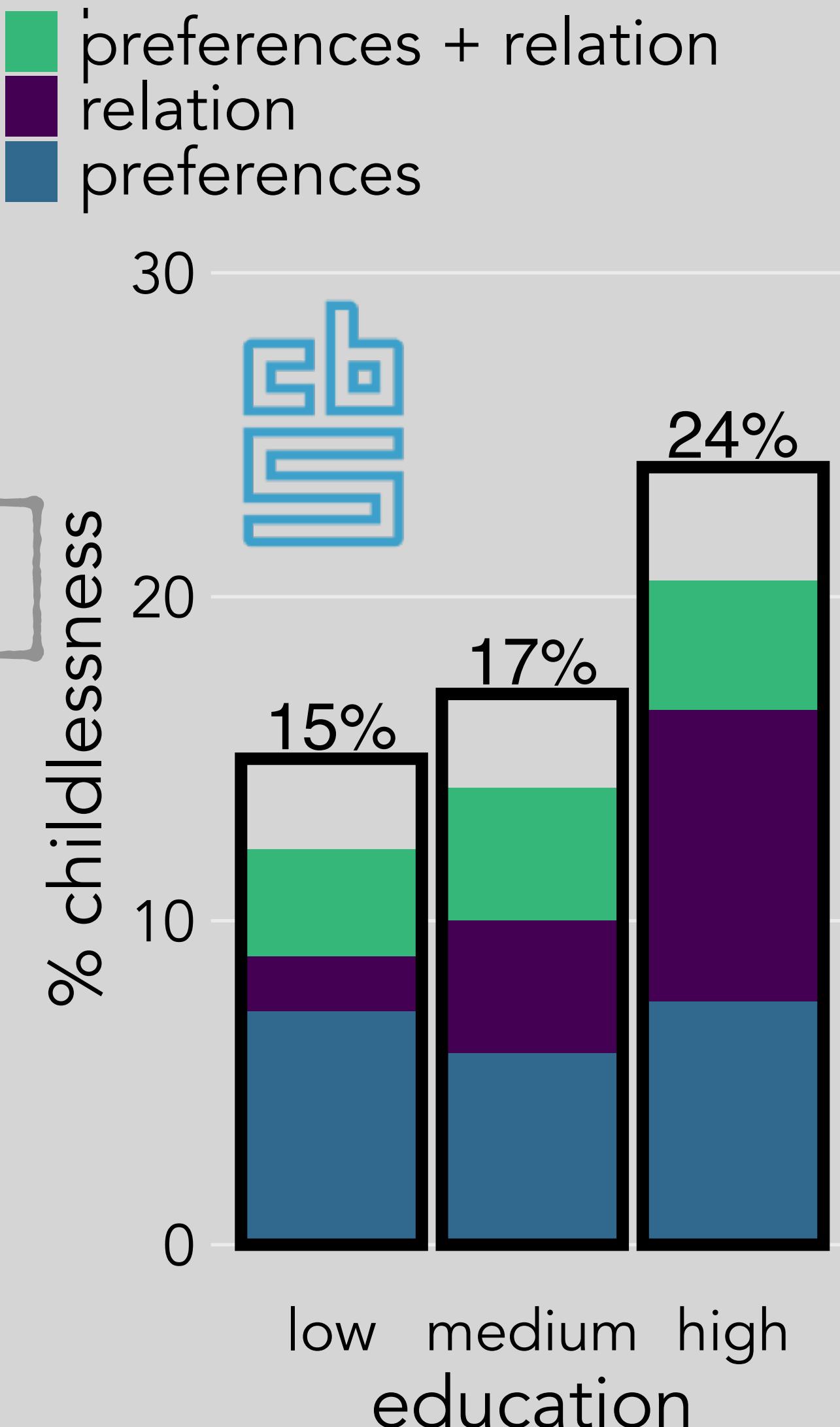




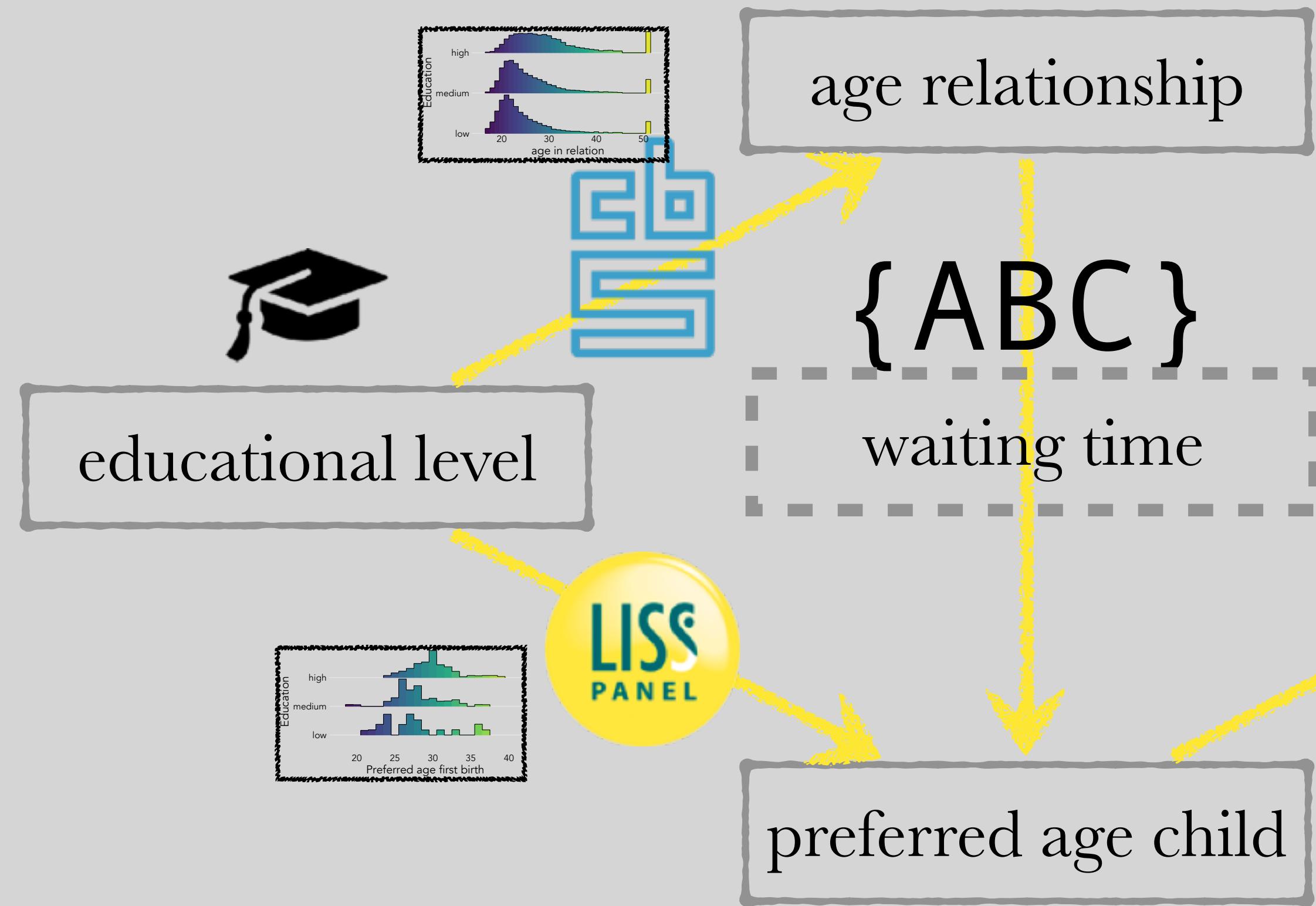


 preferences + relation
 relation
 preferences





{ABC}
Approximate
Bayesian
Computation



{ABC}

age in relation \propto

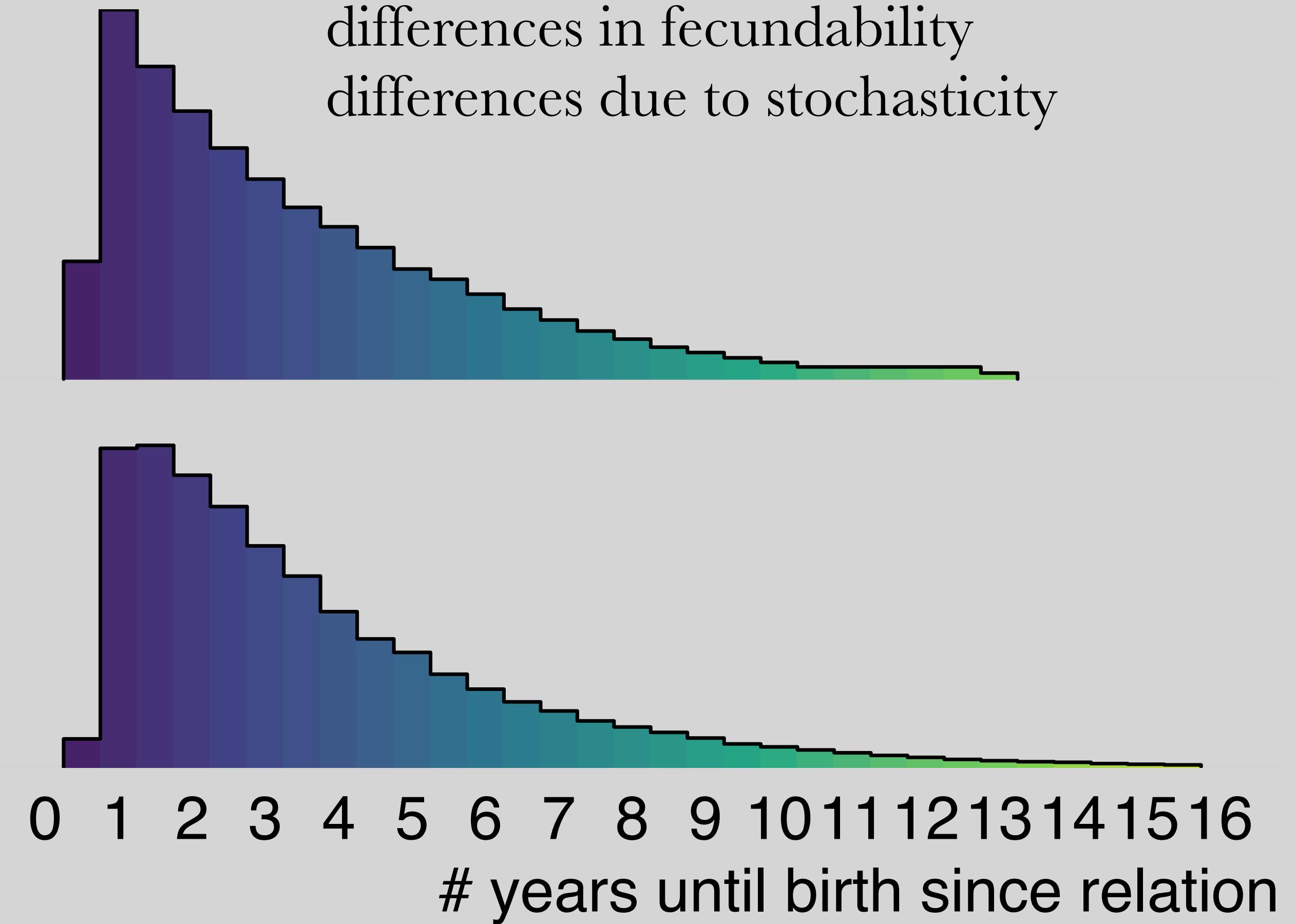
age in relation
fecundability
stochasticity \propto

actual
outcomes

ABC
model

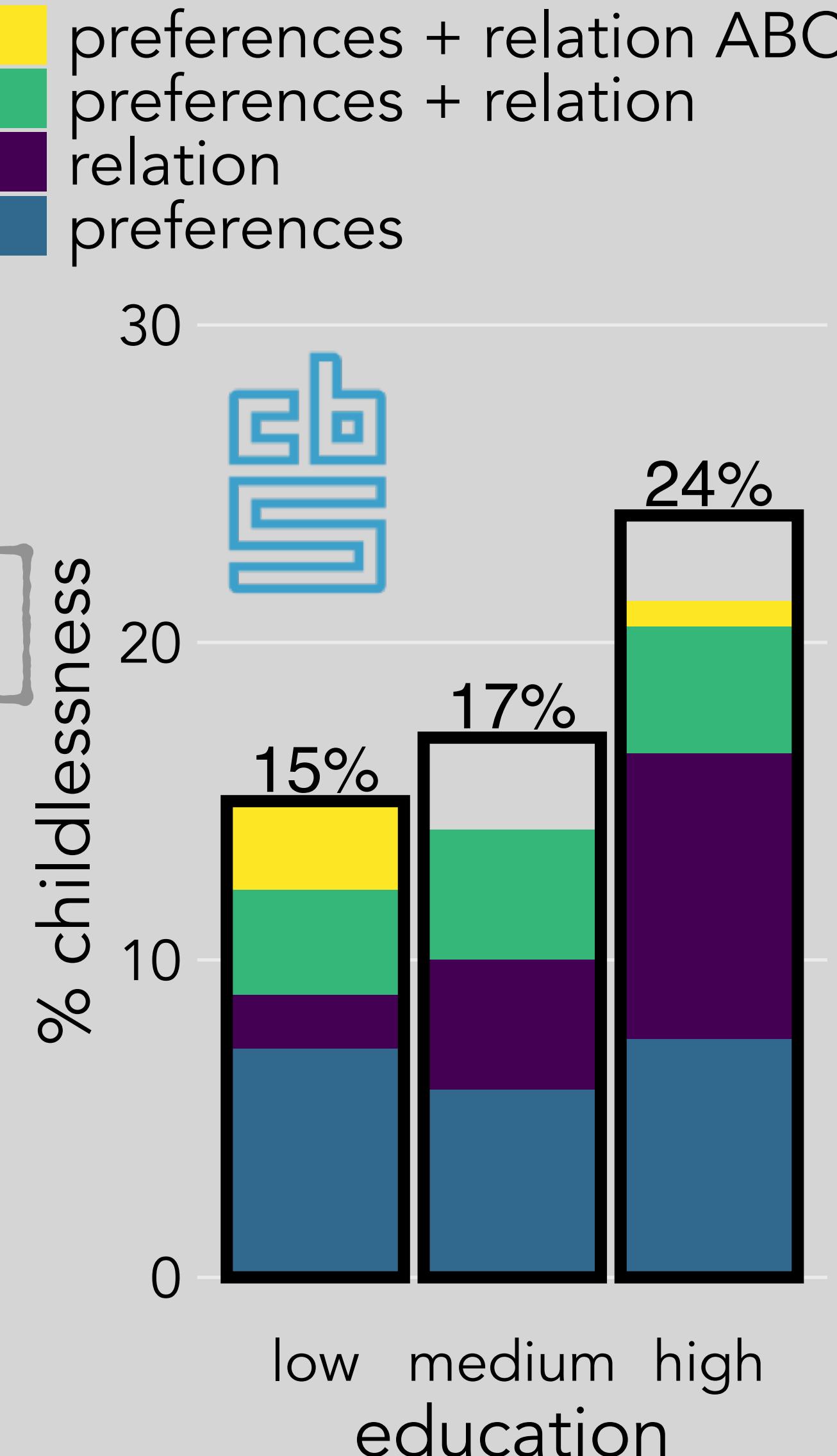
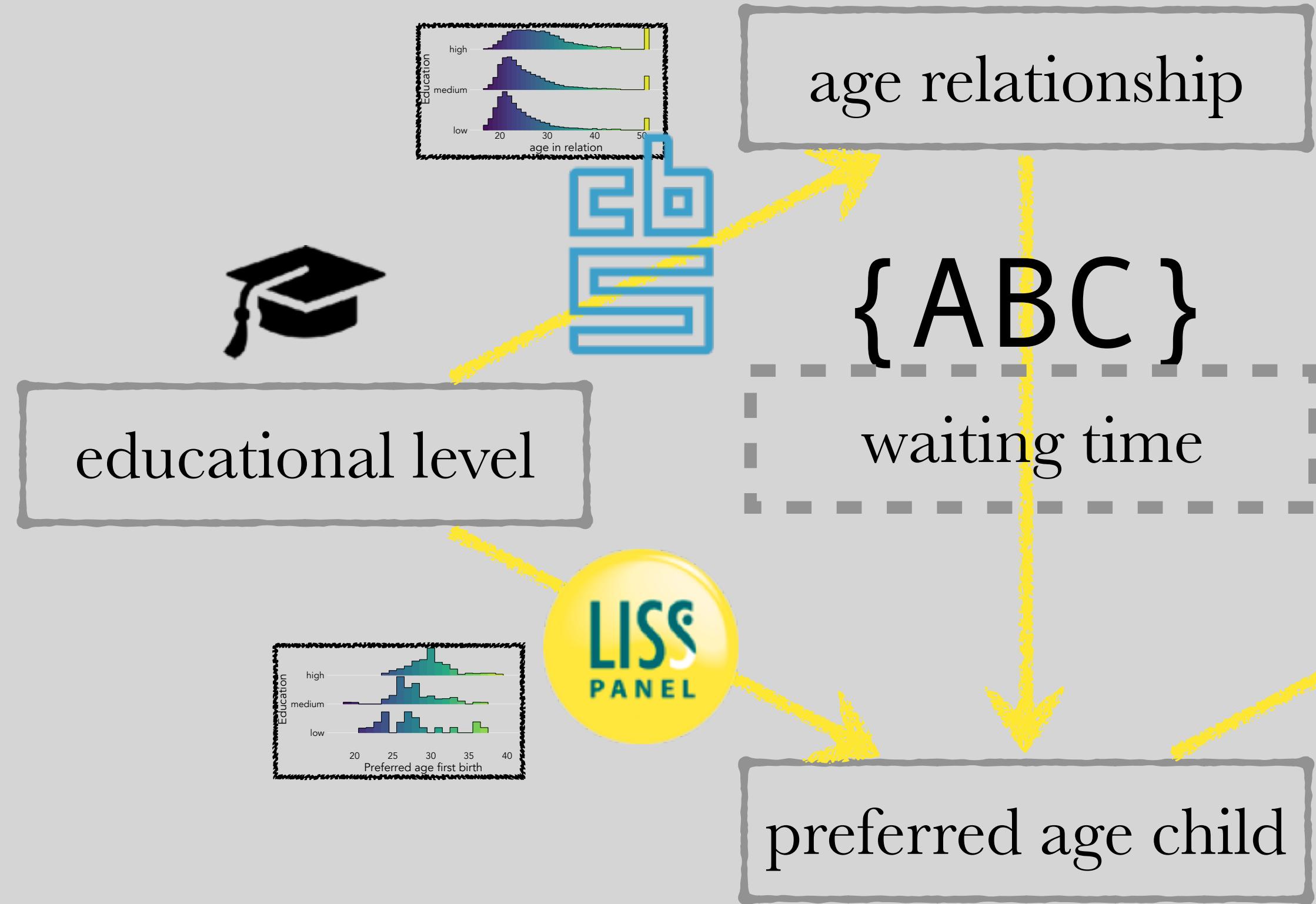
Variation due to:

preferred waiting time child
differences in fecundability
differences due to stochasticity



{ABC}

Approximate Bayesian Computation



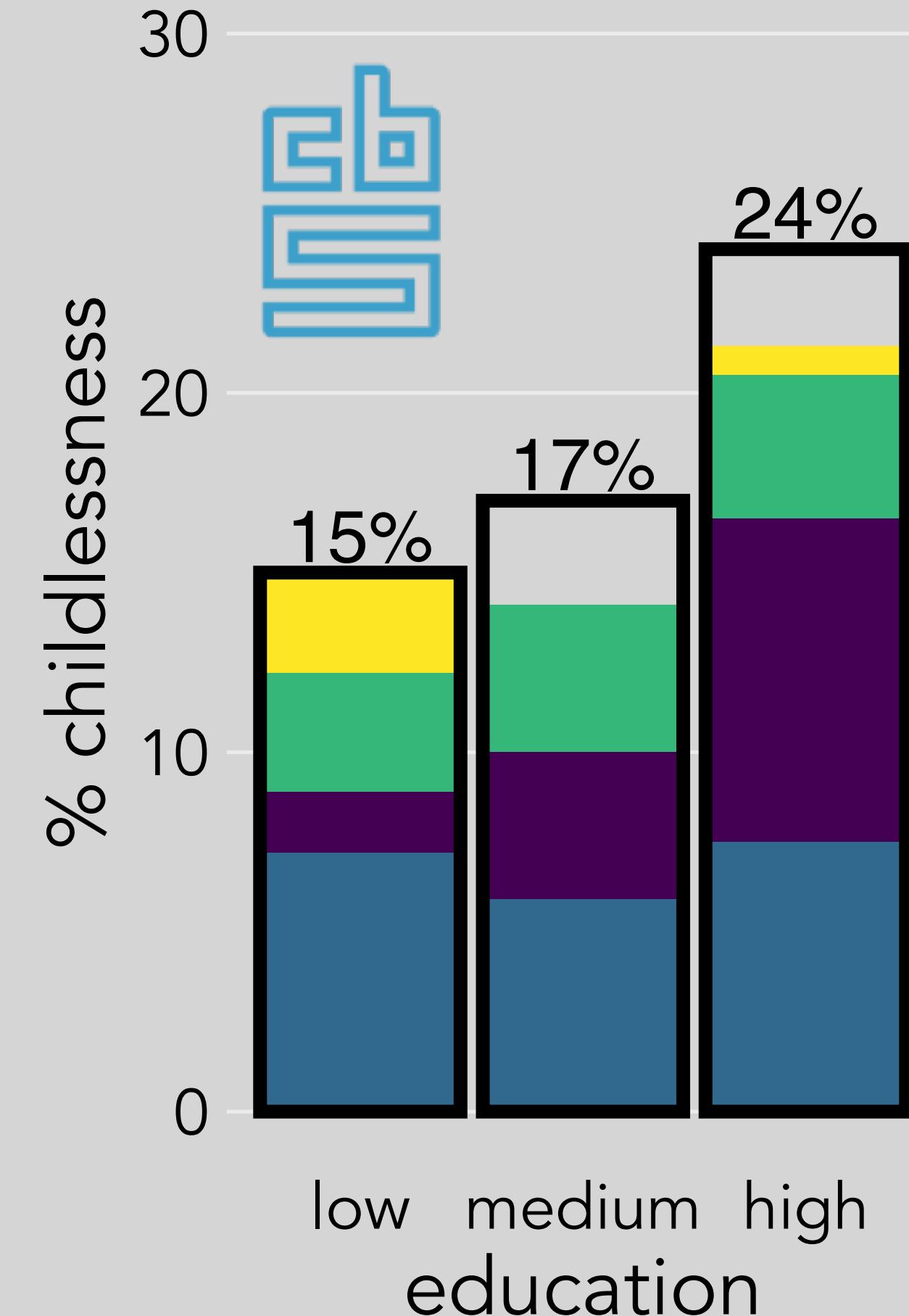
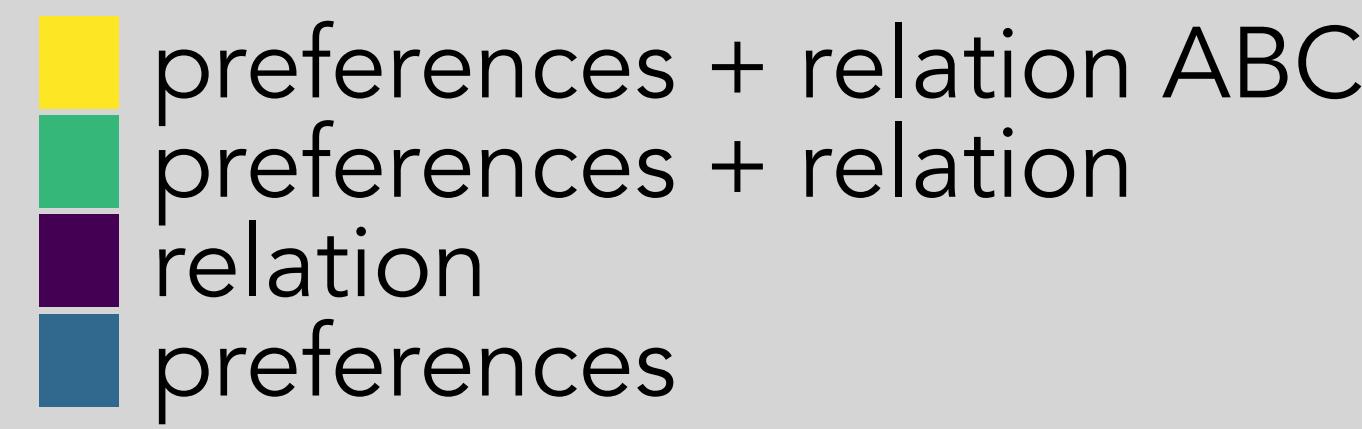
Where Did We Go Wrong?

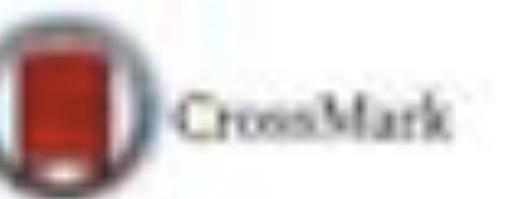
Assumptions

1. No break-ups
2. All births are preferred
3. Preferences do not determine relationship
4. Preferences do not determine education
5. Preferences do not change
6. Education is not related to 'biology'
7. Preferences are measured well

Improvements

1. Make waiting time dependent on age and education
2. Better measures of age in relationship

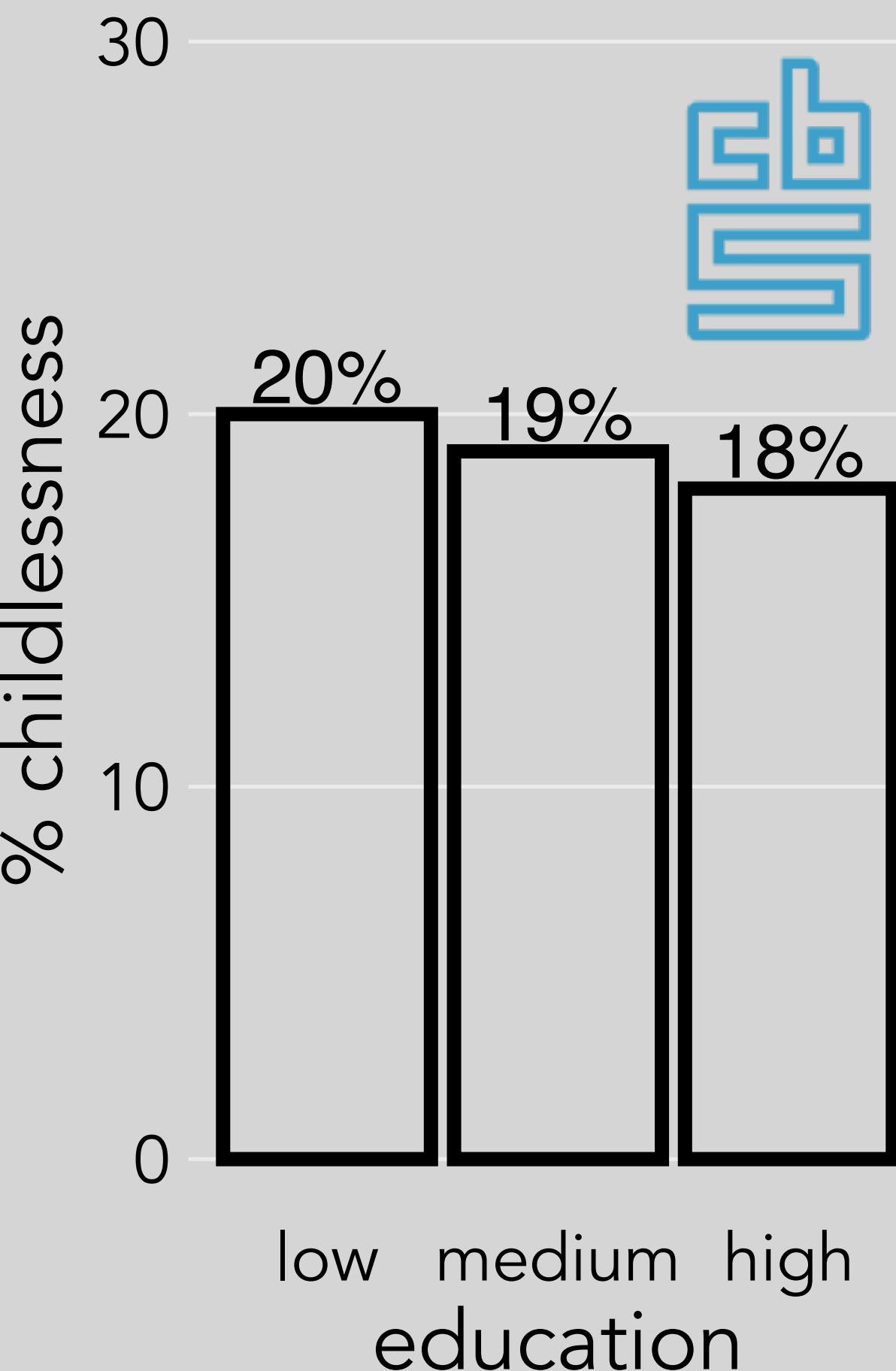




Education, Gender, and Cohort Fertility in the Nordic Countries

Marika Jalovaara¹ • Gerda Neyer² • Gunnar Andersson² • Johan Dahlberg² •
Lars Dommermuth³ • Peter Fallesen^{2,4} • Trude Lappégård⁵

cohort '80-'84



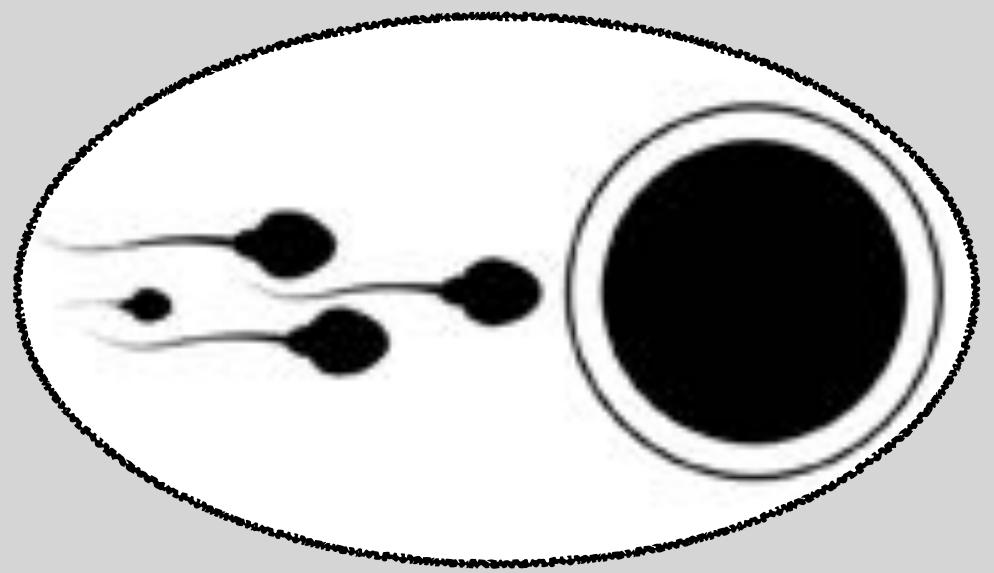
“

In Denmark, Norway and Sweden, childlessness is now highest among the least educated women

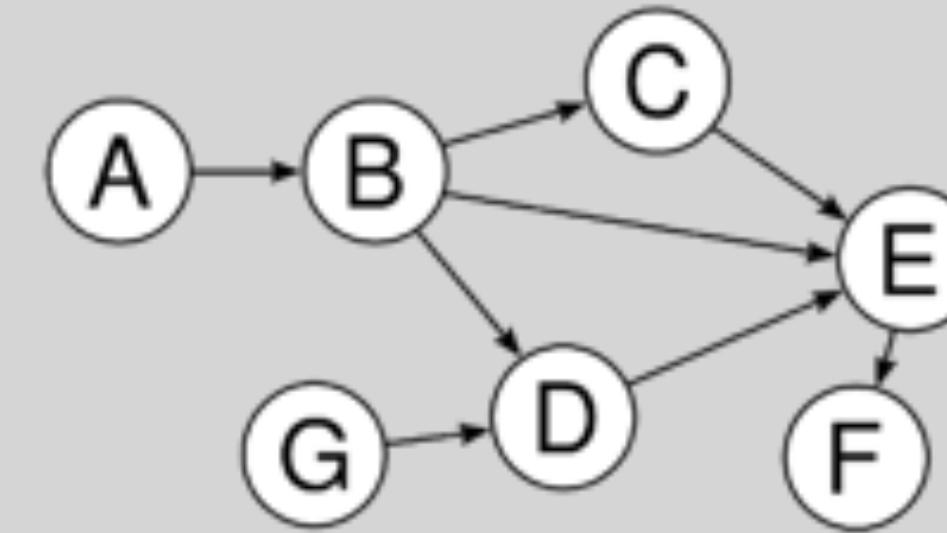
Take-Home Messages

microsimulation can
advance traditional
statistical modelling

microsimulation can:



include
biological
information



test (causal)
mechanisms



quantify
unpredictability



Variation due to Stochasticity ($sd = 13$ months)

Unpredictable Variation!

%

20

15

10

5

0

0

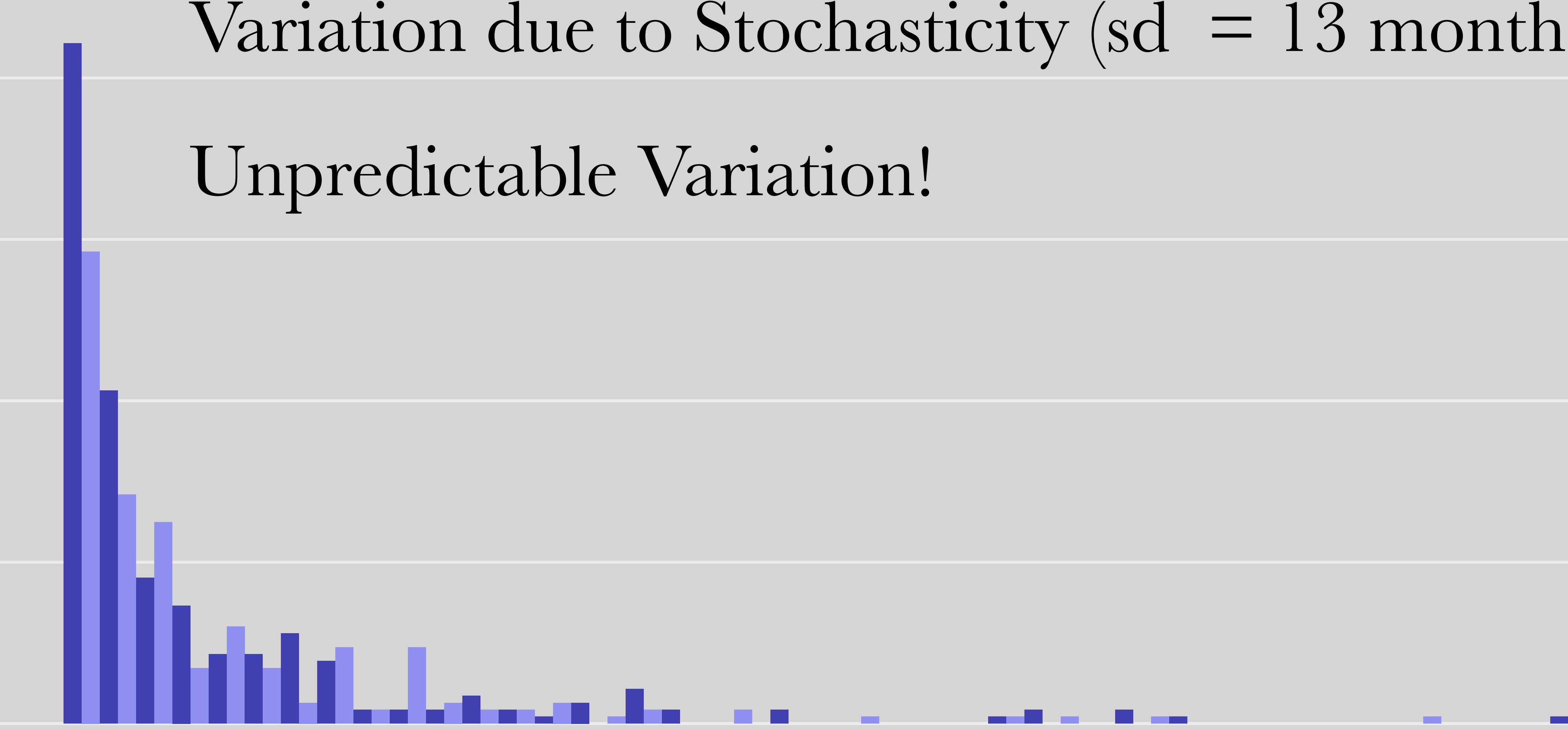
20

40

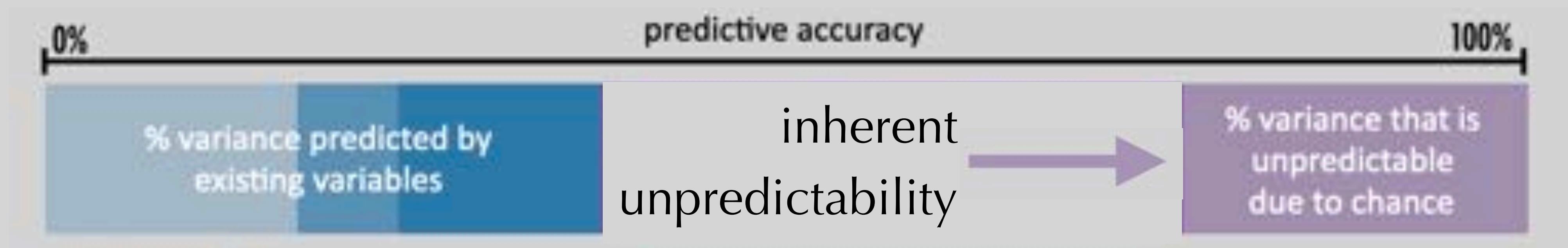
60

80

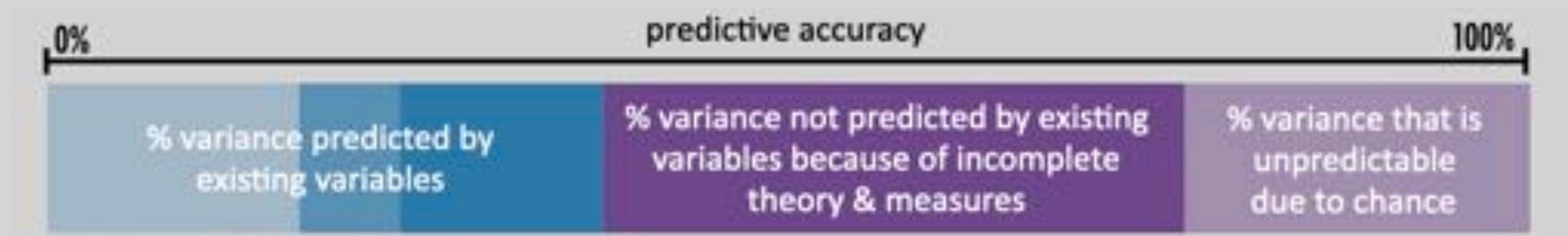
Months until conception for 30 year old women



Unpredictable Variation



Unique Insight into State of Field



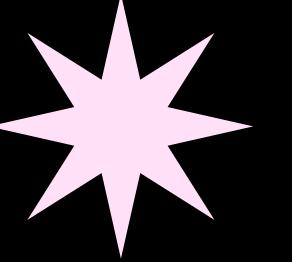
?

The Proposal

a shift towards **prediction**
leads to a more reliable
and useful social science

microsimulation can
advance traditional
statistical modelling

FERTILITY PREDICTION CHALLENGE



⌚ March-August 2024

University of Groningen,
Netherlands

0.54*

Is the current best [known to us] F1-score
of a classifier that predicts who is going
to have a child in the next three years

CAN YOU BEAT THIS SCORE?

Do you want to contribute to research on fertility behavior and the methodology of using prediction in social sciences?

Are you interested in working with unique registry-based datasets, including a social network for the entire Dutch population?

Are you looking for an engaging practical task for your machine learning course or workshop?

Or are you simply curious about the challenge and want to learn more about its design and prizes?



←
Sign up here to receive an update when the registration for the challenge opens and details are available

Contacts:

Gert Stulp g.stulp@rug.nl
Elizaveta Sivak e.sivak@rug.nl



university of
groningen



ODISSEI



Eyra

