This mess we're in?

Or how simulation and prediction will advance fertility research

gert stulp — gertstulp.com — g.stulp@rug.nl

Understanding fertility outcomes by quantifing the (un)predictable

How Well Are We Doing?

variables
explain
little

Minder geboorten door studie en flexwerk?

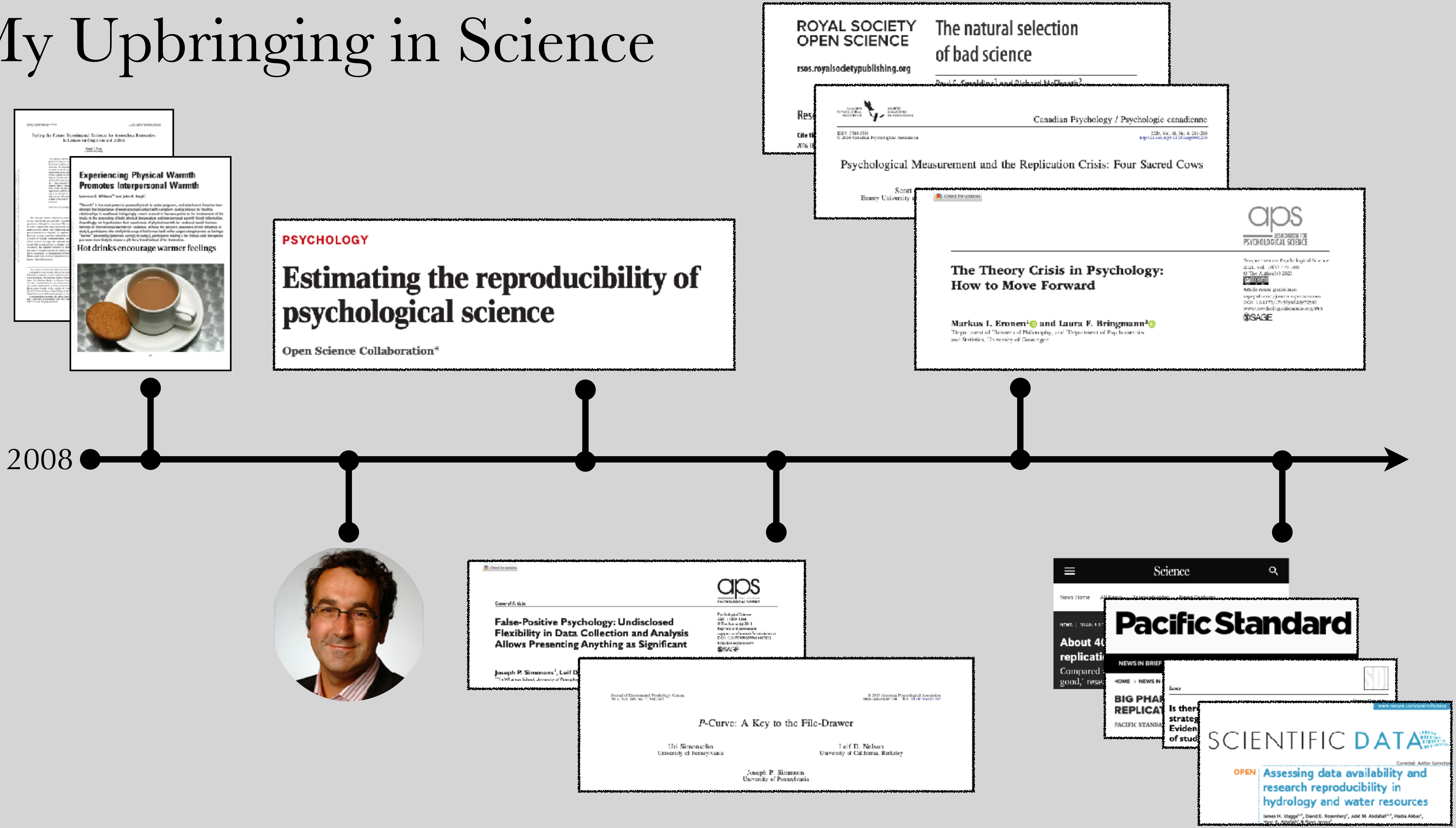
“ total effect on fertility … rather small

surprising
patterns

incomparable
results

non-replicable
results

My Upbringing in Science



Replication (crisis) in Demography?

Reasons why not

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

Reasons why

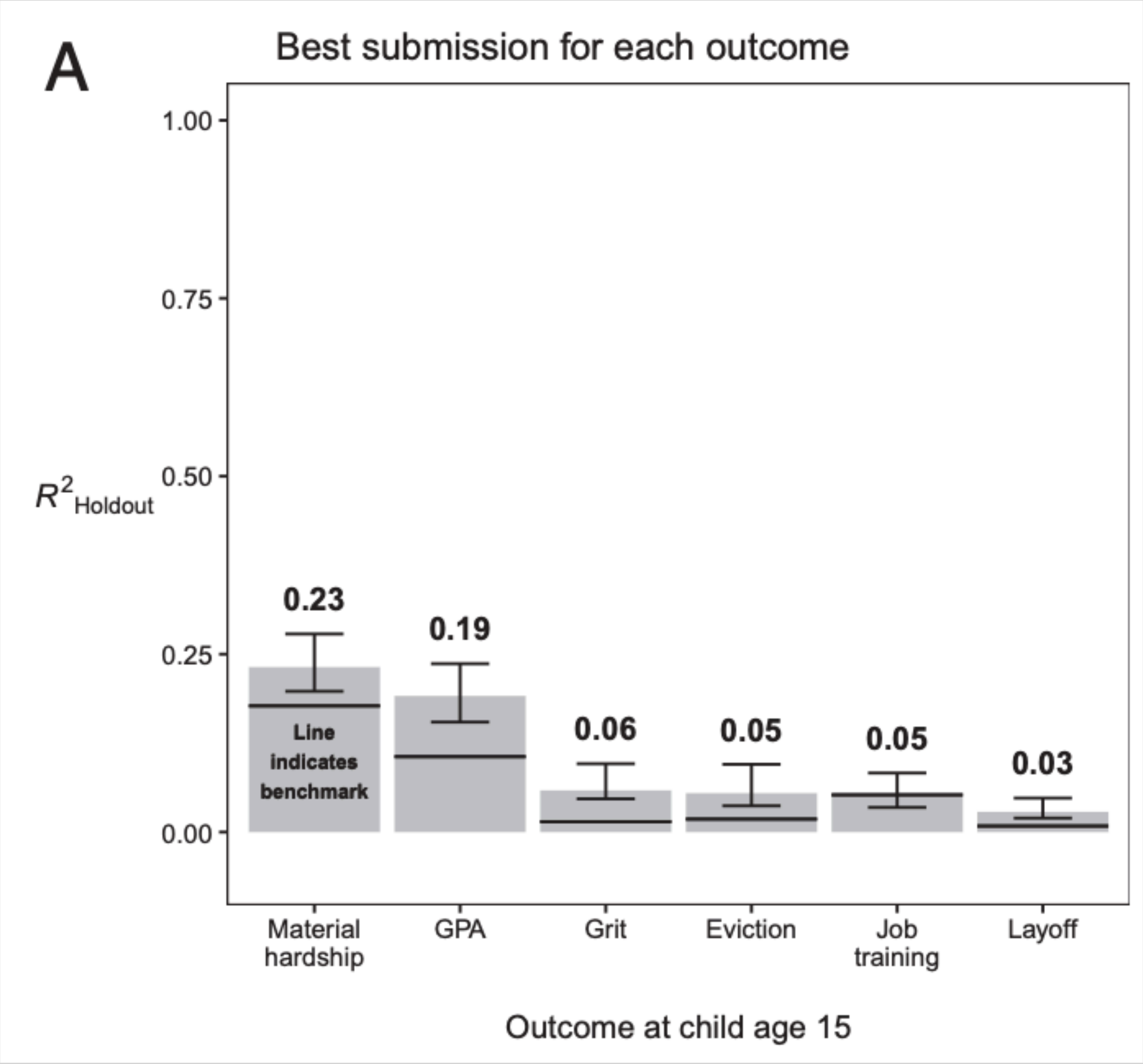
- *Non-experimental*
- *Correlational, but little causal inference*
- *Large N, yet star gazing*
- *Controlling at will*
- *"Culture" as a get-out-of-jail-for-free card*

Predictability Crisis?

Measuring the predictability of life outcomes with a scientific mass collaboration

Matthew J. Salganik[a,1], Ian Lundberg[a], Alexander T. Kindel[a], Caitlin E. Ahearn[b], Khaled Al-Ghoneim[c], Abdullah Almaatouq[d,e], Drew M. Altschul[f], Jennie E. Brand[b,g], Nicole Bohme Carnegie[h], Ryan James Compton[i], Debanjan Datta[j], Thomas Davidson[k], Anna Filippova[l], Connor Gilroy[m], Brian J. Goode[n], Eaman Jahani[o], Ridhi Kashyap[p,q,r], Antje Kirchner[s], Stephen McKay[t], Allison C. Morgan[u], Alex Pentland[e], Kivan Polimis[v], Louis Raes[w], Daniel E. Rigobon[x], Claudia V. Roberts[y], Diana M. Stanescu[z], Yoshihiko Suhara[e], Adaner Usmani[aa], Erik H. Wang[z], Muna Adem[bb], Abdulla Alhajri[cc], Bedoor AlShebli[dd], Redwane Amin[ee], Ryan B. Amos[y], Lisa P. Argyle[ff], Livia Baer-Bositis[gg], Moritz Büchi[hh], Bo-Ryehn Chung[ii], William Eggert[jj], Gregory Faletto[kk], Zhilin Fan[ll], Jeremy Freese[gg], Tejomay Gadgil[mm], Josh Gagné[gg], Yue Gao[nn], Andrew Halpern-Manners[bb], Sonia P. Hashim[y], Sonia Hausen[gg], Guanhua He[oo], Kimberly Higuera[gg], Bernie Hogan[pp], Ilana M. Horwitz[qq], Lisa M. Hummel[gg], Naman Jain[x], Kun Jin[rr], David Jurgens[ss], Patrick Kaminski[bb,tt], Areg Karapetyan[uu,vv], E. H. Kim[gg], Ben Leizman[y], Naijia Liu[z], Malte Möser[y], Andrew E. Mack[z], Mayank Mahajan[y], Noah Mandell[ww], Helge Marahrens[bb], Diana Mercado-Garcia[qq], Viola Mocz[xx], Katariina Mueller-Gastell[gg], Ahmed Musse[yy], Qiankun Niu[ee], William Nowak[zz], Hamidreza Omidvar[aaa], Andrew Or[y], Karen Ouyang[y], Katy M. Pinto[bbb], Ethan Porter[ccc], Kristin E. Porter[ddd], Crystal Qian[y], Tamkinat Rauf[gg], Anahit Sargsyan[eee], Thomas Schaffner[y], Landon Schnabel[gg], Bryan Schonfeld[z], Ben Sender[fff], Jonathan D. Tang[y], Emma Tsurkov[gg], Austin van Loon[gg], Onur Varol[ggg,hhh], Xiafei Wang[iii], Zhi Wang[hhh,jjj], Julia Wang[y], Flora Wang[fff], Samantha Weissman[y], Kirstie Whitaker[kkk,lll], Maria K. Wolters[mmm], Wei Lee Woon[nnn], James Wu[ooo], Catherine Wu[y], Kengran Yang[aaa], Jingwen Yin[ll], Bingyu Zhao[ppp], Chenyun Zhu[ll], Jeanne Brooks-Gunn[qqq,rrr], Barbara E. Engelhardt[y,ii], Moritz Hardt[sss], Dean Knox[z], Karen Levy[ttt], Arvind Narayanan[y], Brandon M. Stewart[a], Duncan J. Watts[uuu,vvv,www], and Sara McLanahan[a,1]

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

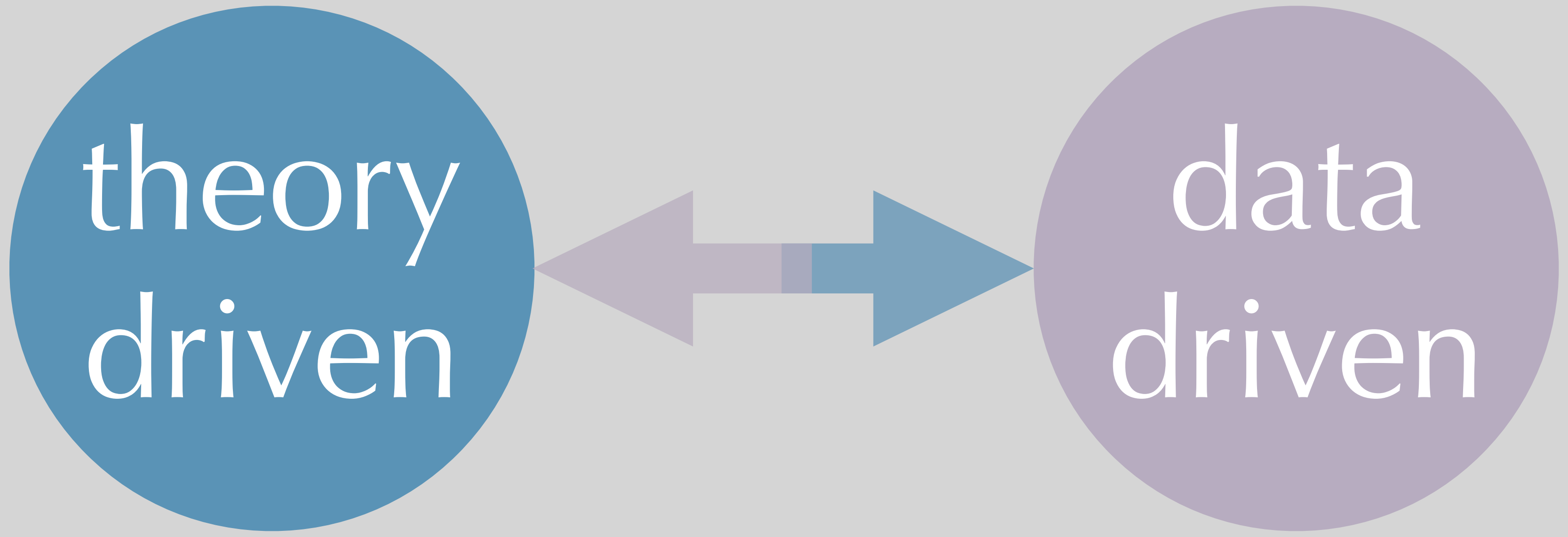
microsimulation can advance traditional statistical modelling

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

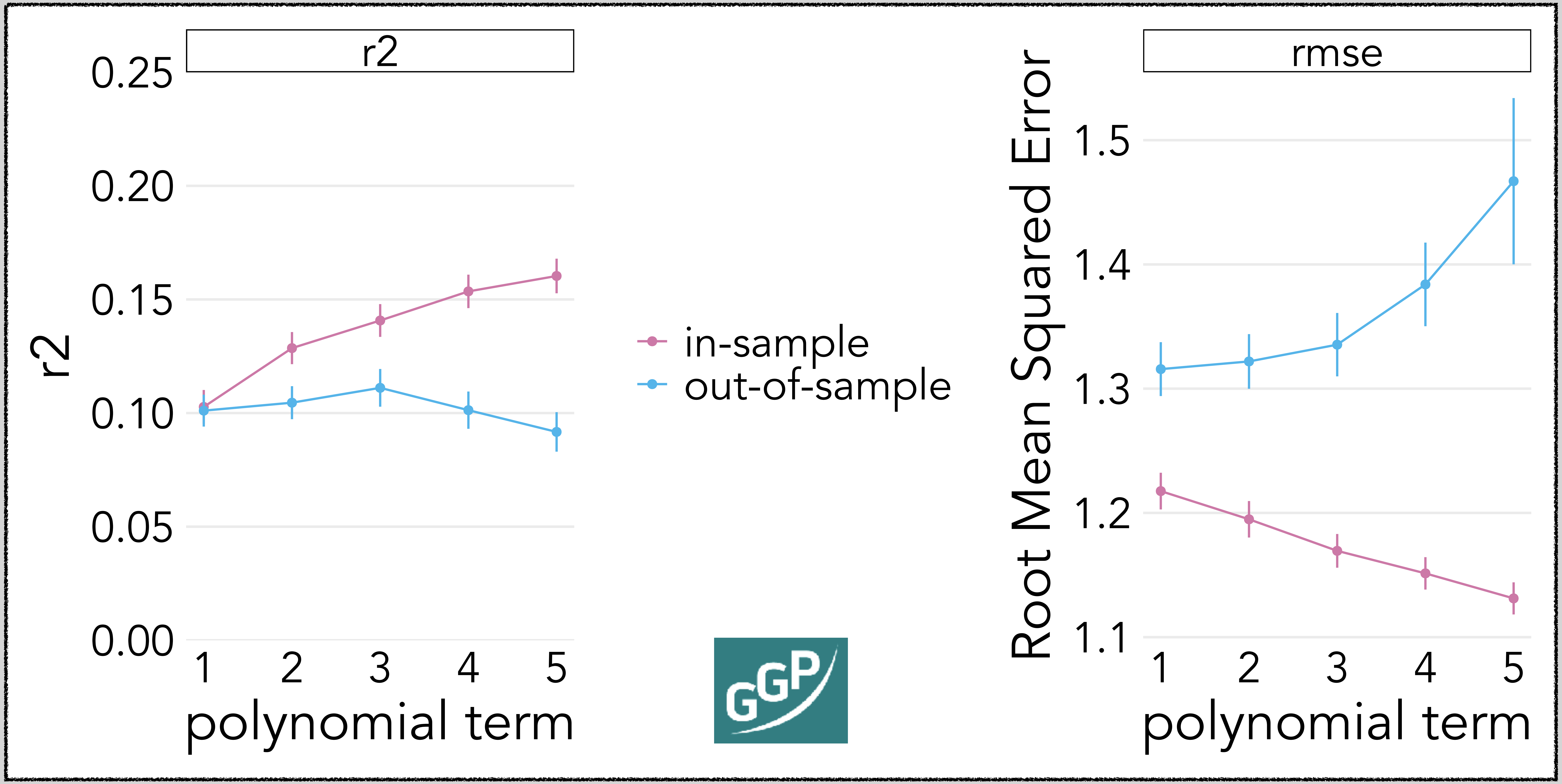


facilitates dialogue theory- and data-driven models

measure of distance theory and practice

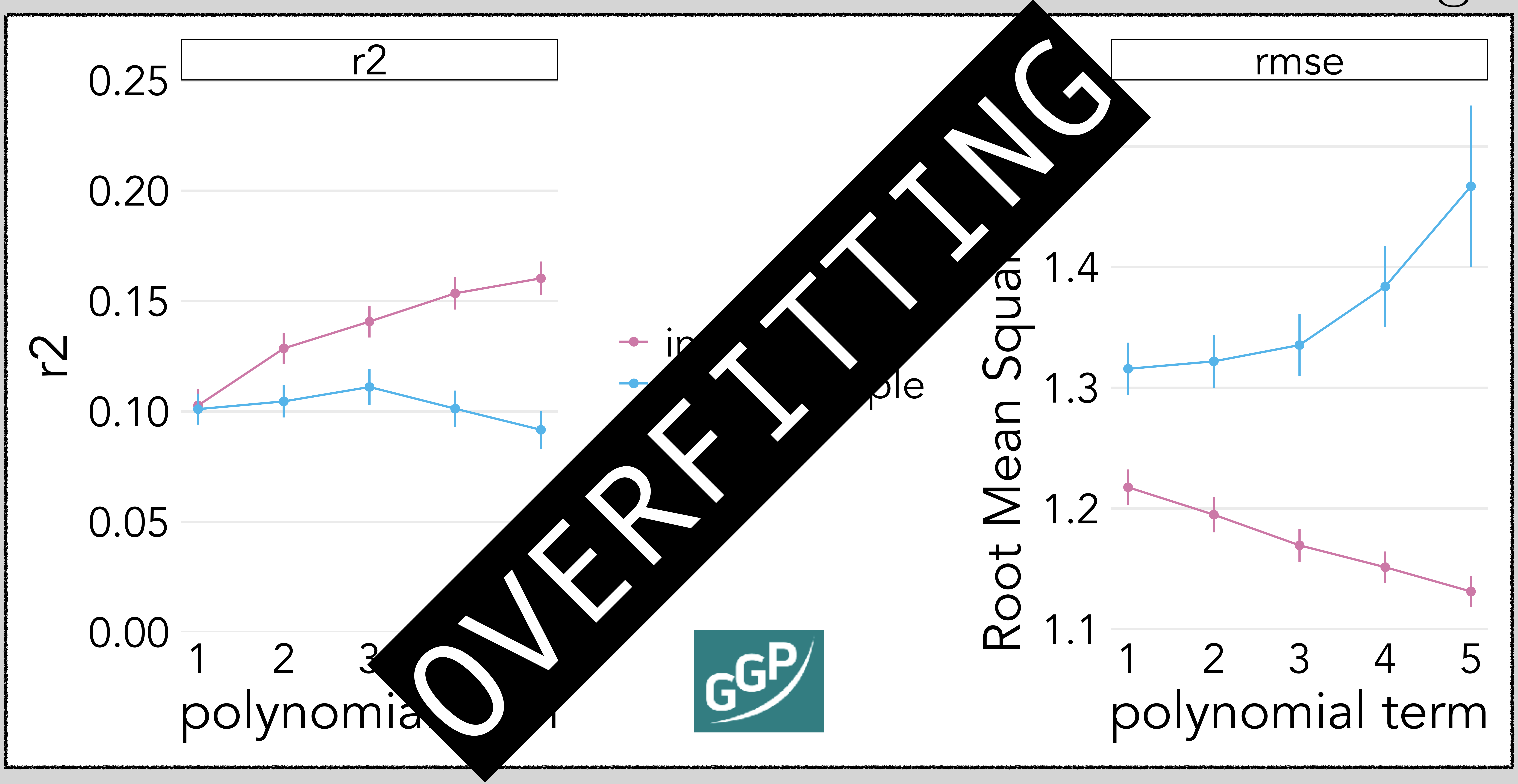
Out-of-Sample Prediction

$n = 50$ training data
$n = 50$ test data
number of children ~ age



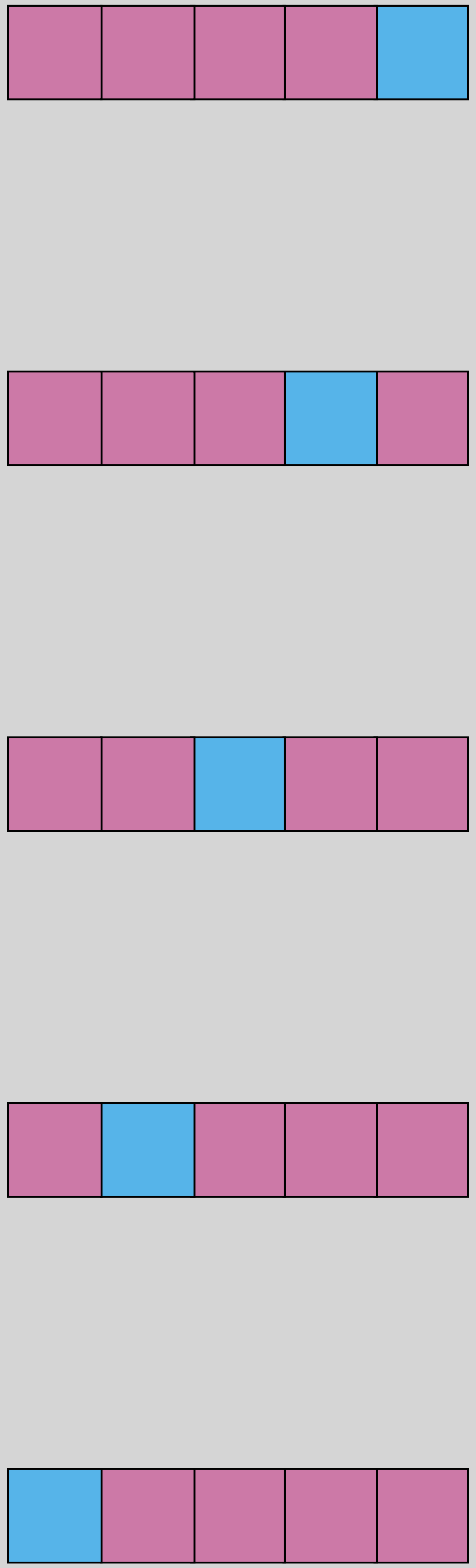
Out-of-Sample Prediction

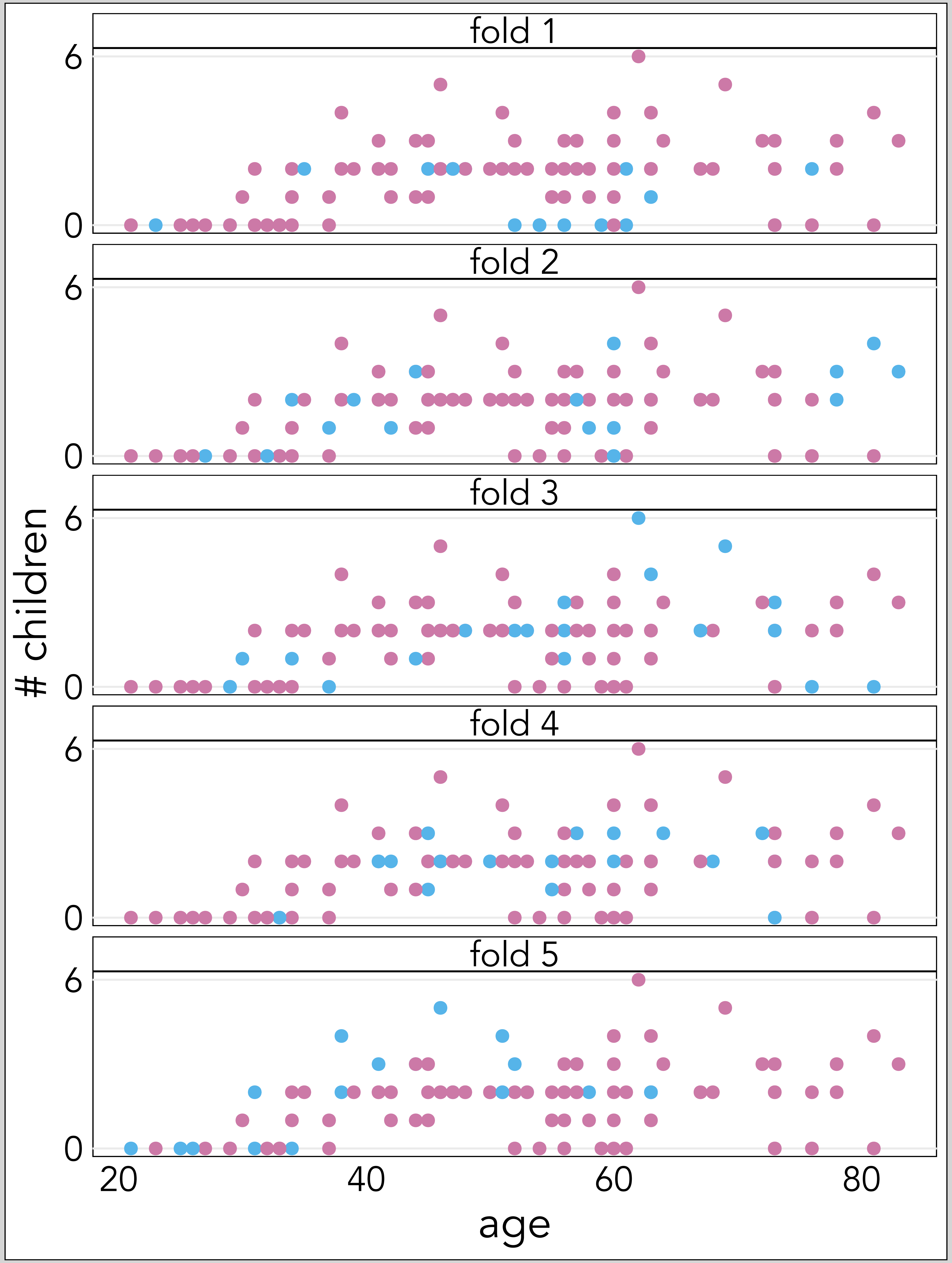
n = 50 *training data*
n = 50 *test data*
number of children ~ age



Cross Validation

fold	in-sample R^2	out-of-sample R^2
1	0.15	0.12
2	0.17	0.17
3	0.14	0.20
4	0.20	-0.18
5	0.27	-1.38
Average out-of-sample R^2		-0.07



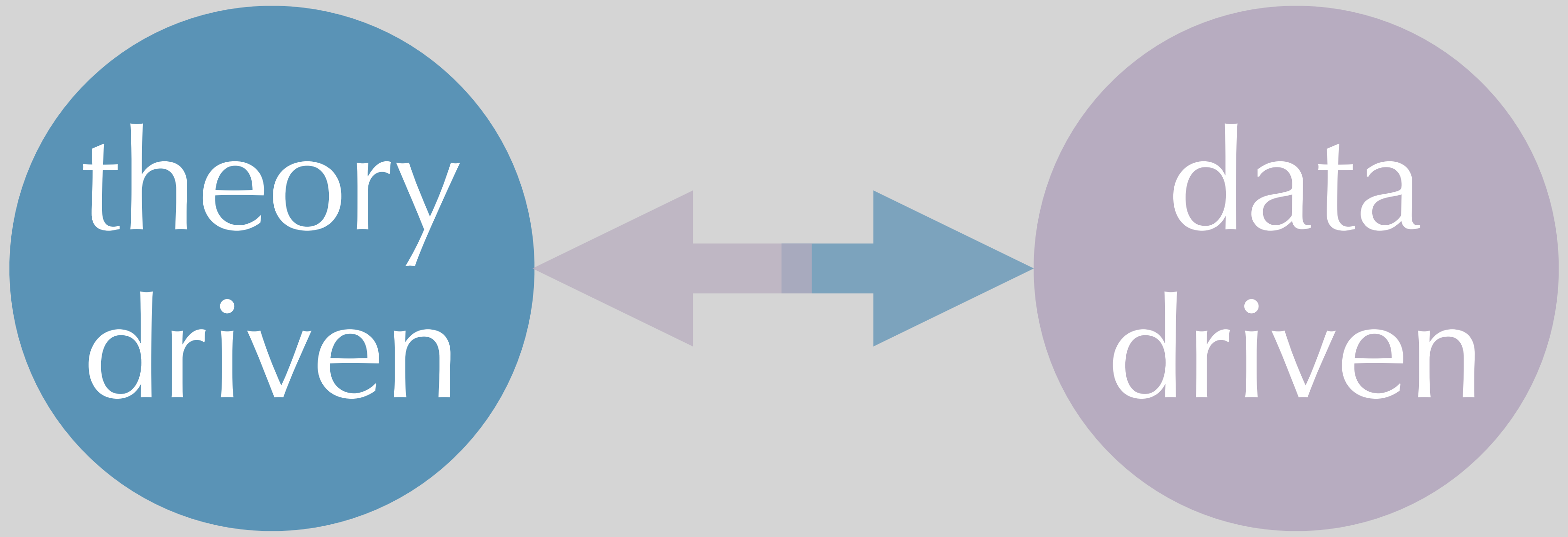


The Proposal

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

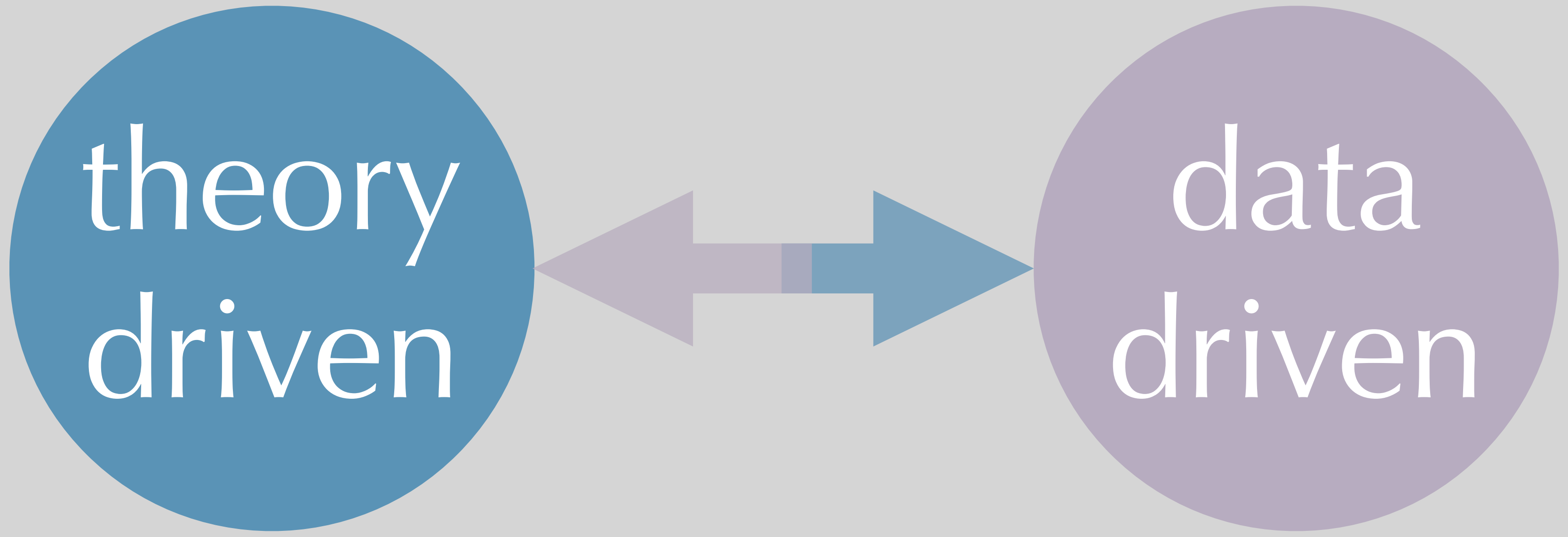
- *is easy(ier) to understand*
- *can be compared across analytical techniques*
- *can be compared across models*
- *is less gameable*

The Proposal

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

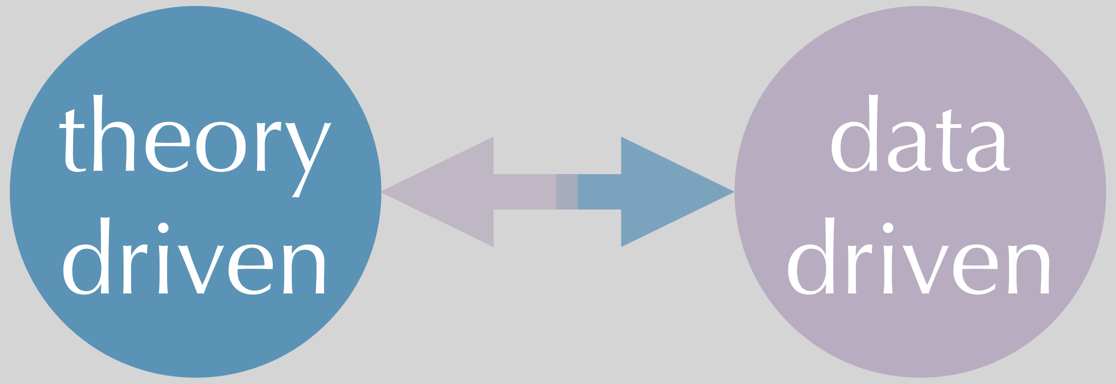
out-of-sample predictive ability:

clear measure of effect size



facilitates dialogue theory- and data-driven models

measure of distance theory and practice



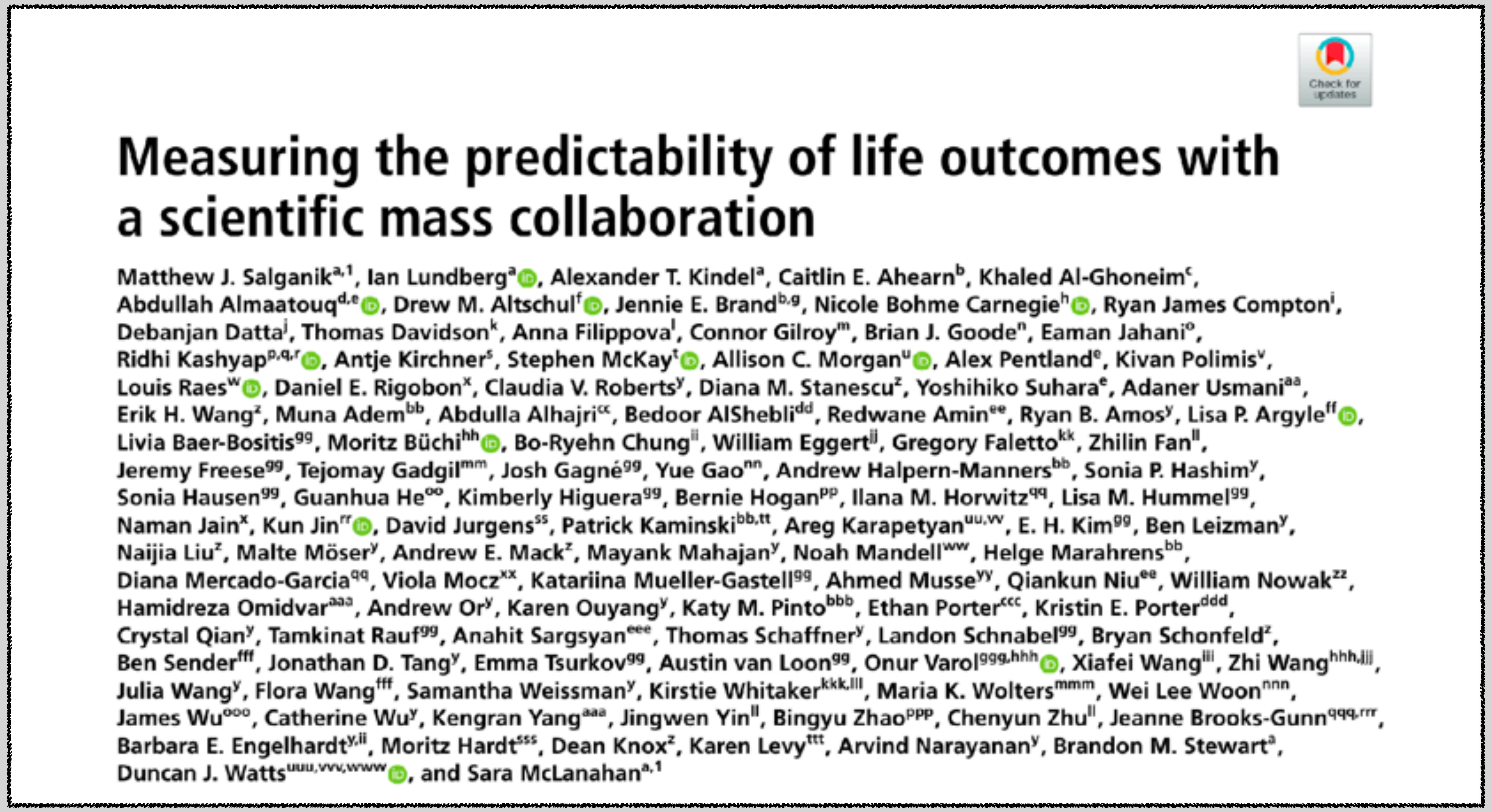
theory-driven *vs* data-driven

theory-driven	data-driven
focus on (causal) estimates	*focus on predictive ability*
support based on p-value	*support based on prediction*
limited number of variables (k)	*k may be larger than n*

NHST weird theory-testing
long reign the linear model
pet variable problem

estimates uninterpretable (sort of)
computing intensive

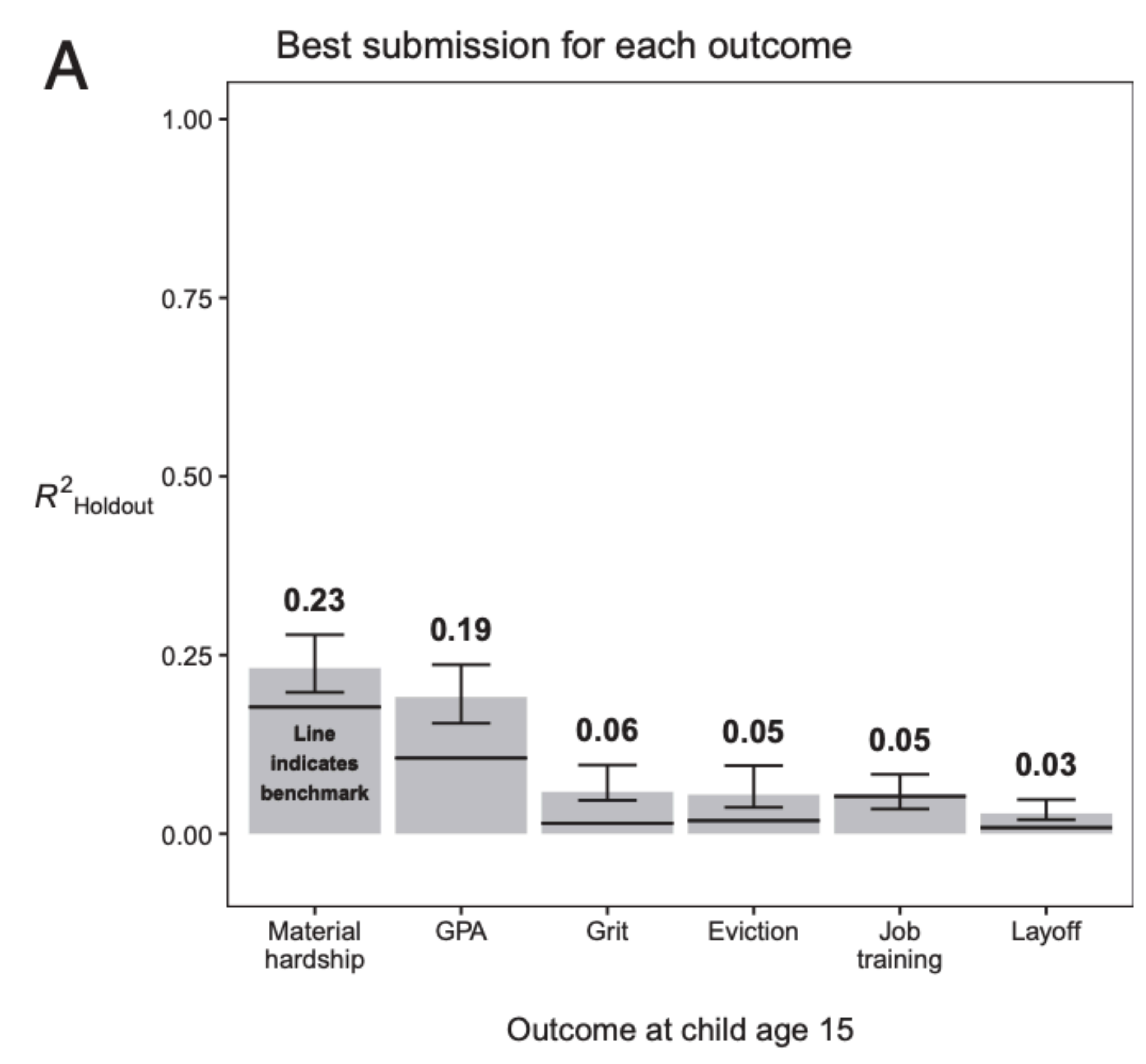
Predictability Crisis?


Check for updates

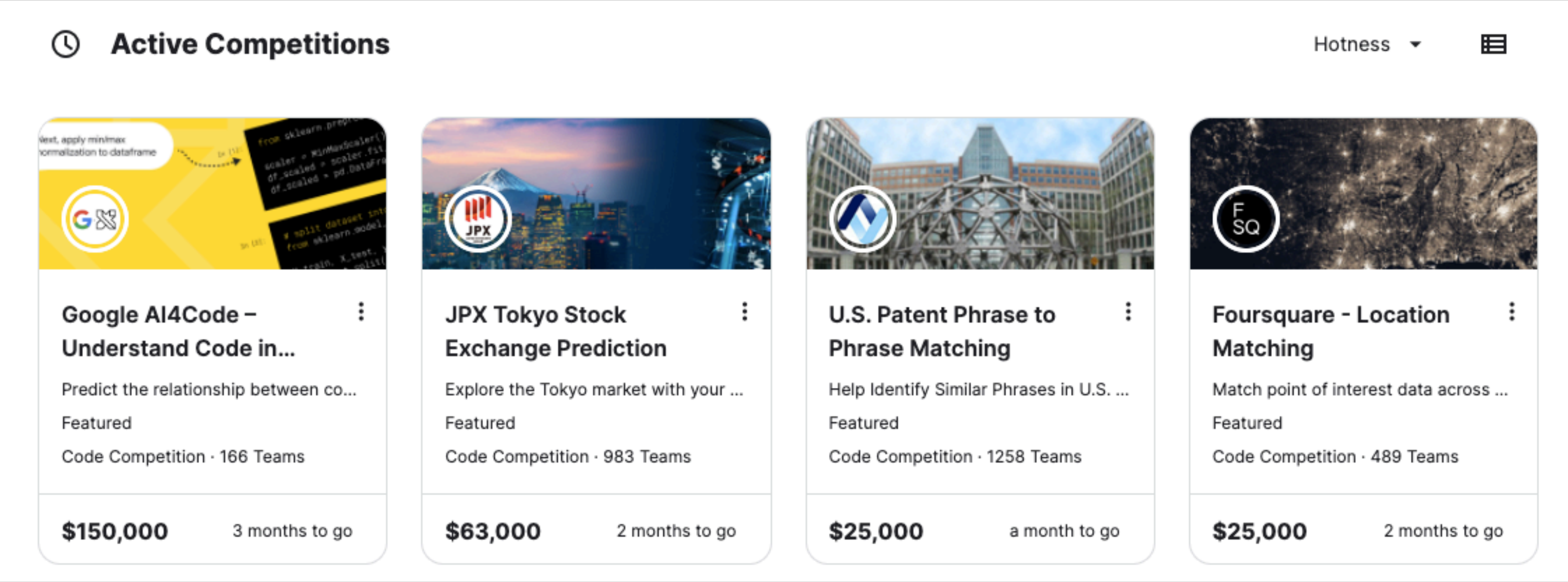
Measuring the predictability of life outcomes with a scientific mass collaboration

Matthew J. Salganik[a,1], Ian Lundberg[a], Alexander T. Kindel[a], Caitlin E. Ahearn[b], Khaled Al-Ghoneim[c], Abdullah Almaatouq[d,e], Drew M. Altschul[f], Jennie E. Brand[b,g], Nicole Bohme Carnegie[h], Ryan James Compton[i], Debanjan Datta[j], Thomas Davidson[k], Anna Filippova[l], Connor Gilroy[m], Brian J. Goode[n], Eaman Jahani[o], Ridhi Kashyap[p,q,r], Antje Kirchner[s], Stephen McKay[t], Allison C. Morgan[u], Alex Pentland[e], Kivan Polimis[v], Louis Raes[w], Daniel E. Rigobon[x], Claudia V. Roberts[y], Diana M. Stanescu[z], Yoshihiko Suhara[e], Adaner Usmani[aa], Erik H. Wang[z], Muna Adem[bb], Abdulla Alhajri[cc], Bedoor AlShebli[dd], Redwane Amin[ee], Ryan B. Amos[y], Lisa P. Argyle[ff], Livia Baer-Bositis[gg], Moritz Büchi[hh], Bo-Ryehn Chung[ii], William Eggert[jj], Gregory Faletto[kk], Zhilin Fan[ll], Jeremy Freese[gg], Tejomay Gadgil[mm], Josh Gagné[gg], Yue Gao[nn], Andrew Halpern-Manners[bb], Sonia P. Hashim[y], Sonia Hausen[gg], Guanhua He[oo], Kimberly Higuera[gg], Bernie Hogan[pp], Ilana M. Horwitz[qq], Lisa M. Hummel[gg], Naman Jain[x], Kun Jin[rr], David Jurgens[ss], Patrick Kaminski[bb,tt], Areg Karapetyan[uu,vv], E. H. Kim[gg], Ben Leizman[y], Naijia Liu[z], Malte Möser[y], Andrew E. Mack[z], Mayank Mahajan[y], Noah Mandell[ww], Helge Marahrens[bb], Diana Mercado-Garcia[qq], Viola Mocz[xx], Katariina Mueller-Gastell[gg], Ahmed Musse[yy], Qiankun Niu[ee], William Nowak[zz], Hamidreza Omidvar[aaa], Andrew Or[y], Karen Ouyang[y], Katy M. Pinto[bbb], Ethan Porter[ccc], Kristin E. Porter[ddd], Crystal Qian[y], Tamkinat Rauf[gg], Anahit Sargsyan[eee], Thomas Schaffner[y], Landon Schnabel[gg], Bryan Schonfeld[z], Ben Sender[fff], Jonathan D. Tang[y], Emma Tsurkov[gg], Austin van Loon[gg], Onur Varol[ggg,hhh], Xiafei Wang[iii], Zhi Wang[hhh,jjj], Julia Wang[y], Flora Wang[fff], Samantha Weissman[y], Kirstie Whitaker[kkk,lll], Maria K. Wolters[mmm], Wei Lee Woon[nnn], James Wu[ooo], Catherine Wu[y], Kengran Yang[aaa], Jingwen Yin[ll], Bingyu Zhao[ppp], Chenyun Zhu[ll], Jeanne Brooks-Gunn[qqq,rrr], Barbara E. Engelhardt[y,ii], Moritz Hardt[sss], Dean Knox[z], Karen Levy[ttt], Arvind Narayanan[y], Brandon M. Stewart[a], Duncan J. Watts[uuu,vvv,www], and Sara McLanahan[a,1]

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



kaggle



“secret sauce of data science

Donoho, 2015

data
challenge

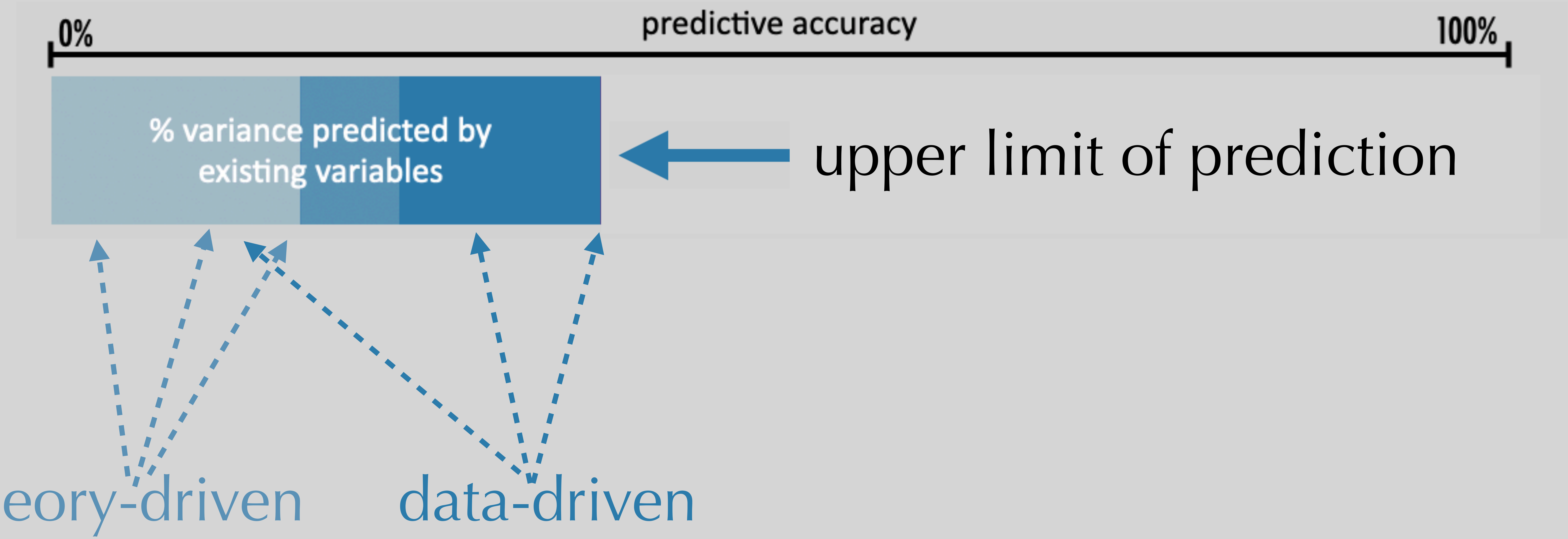
theory driven

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

baseline benchmarks	upper limit benchmarks
established with state-of-the-art theory	*established with state-of-the-art statistical learning tools*

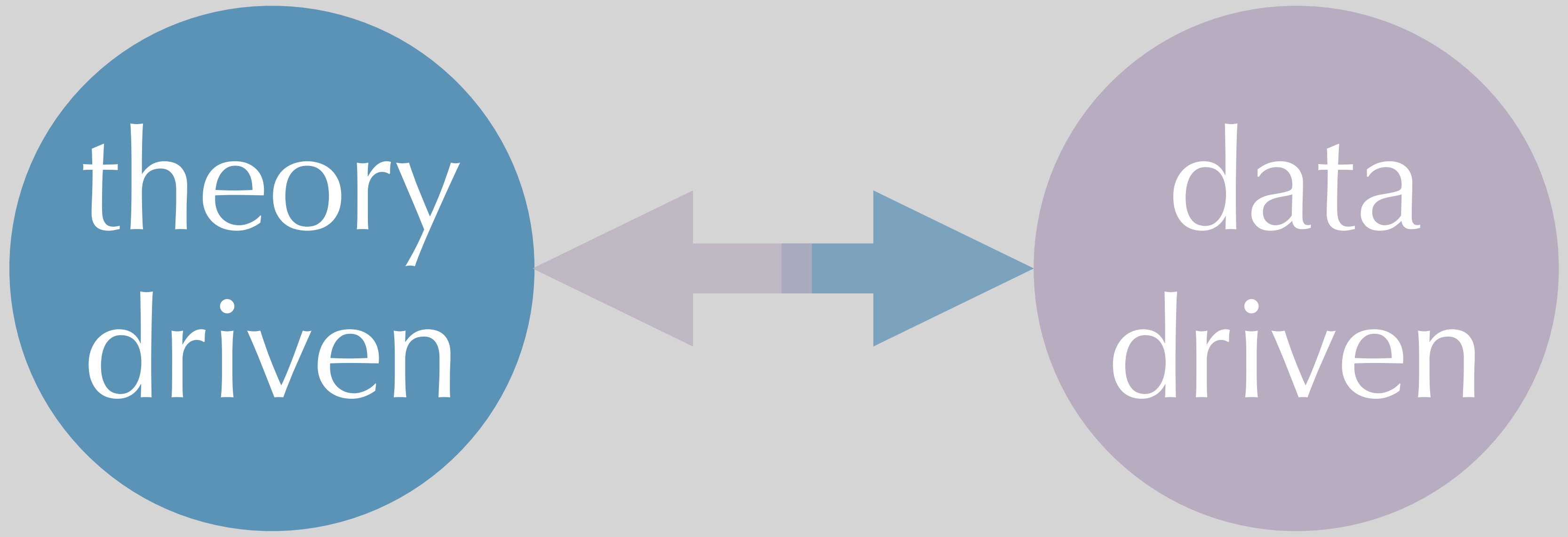
> Progress usually comes from many small improvements; a change of 1% can be a reason to break out the champagne
>
> Liberman, 2012

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 a measure of how useful
our theory is in the real world

Journal of Peace Research

Articles

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

Journal of Peace Research
47(4) 363–375

Reprints and permission:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/0022343309356491

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

Check for updates

aps ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

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

Perspectives on Psychological Science
1–11

Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1745691620969650
www.psychologicalscience.org/PPS
SAGE

Elliot T. Berkman and Sylas M. Wilson
Department of Psychology and Center for Translational Neuroscience, University of Oregon

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

PNAS

CrossMark
click for updates

Why significant variables aren't automatically good predictors

Adeline Lo[a], Herman Chernoff[b,1], Tian Zheng[c], and Shaw-Hwa Lo[c,1]

[a]Department of Political Science, University of California, San Diego, La Jolla, CA 92093; [b]Department of Statistics, Harvard University, Cambridge, MA 02138; and [c]Department 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

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.

SOCIUS
Original Article

Socius: Sociological Research for a Dynamic World
Volume 8: 1–17
© The Author(s) 2022
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/23780231221081702
srd.sagepub.com
SAGE

A Pragmatist's Guide to Using Prediction in the Social Sciences

Mark D. Verhagen[1,2]

Abstract
Prediction is an underused tool in the social sciences, often for the wrong reasons. Many social scientists confuse prediction with unnecessarily complicated methods or with narrowly predicting the future. This is unfortunate. When we view prediction as the simple process of evaluating a model's ability to approximate an outcome of interest, it becomes a more generally applicable and disarmingly simple technique. For all its simplicity, the value of prediction should not be underestimated. Prediction can address enduring sources of criticism plaguing the social sciences, like a lack of assessing a model's ability to reflect the real world, or the use of overly simplistic models to capture social life. The author illustrates these benefits with empirical examples that merely skim the surface of the many and varied ways in which prediction can be applied, staking the claim that prediction is a truly illustrious "free lunch" that can greatly benefit social scientists in their empirical work.

Keywords
prediction, computational social sciences, explanation

Social scientists should start using prediction more often. Prediction is the process of generating predicted values of a dependent variable by applying an estimated model to a set of explanatory variables, bringing a unique analytical perspective to empirical work. Prediction can also help address enduring sources of criticism facing the social sciences. Examples are a general lack of assessing research findings in terms of their real-world relevance, and the use of overly simplistic models to study the complexities of social life. In this article, I address common misconceptions about prediction and provide a simple definition that addresses existing barriers to adoption. I then discuss and illustrate some of the many benefits that prediction can bring when used as a complement to traditional empirical methods. I argue that prediction can and should become a fundamental part of the social scientist's empirical toolkit but that this first requires us to look beyond the current dichotomy between prediction and explanation and instead view the two as complementary to each other.

The current lack of prediction in the social sciences stems from a seeming incompatibility between wanting to explain and wanting to predict, effectively forcing researchers to choose between the two approaches. A case in point is the much-cited article by Galit Shmueli (2010), aptly titled "To Predict or to Explain," which outlines how a social scientist's empirical work flow differs in terms of data processing, modeling, and postestimation diagnostics when choosing to either predict or explain. Naturally, Shmueli assumes that a researcher would not normally attempt to do both. This is an accurate reflection of social science research. The apparent need to dogmatically choose between either approach means that, in practice, social scientists tend to stick to explanation almost exclusively. Illustratively, the terms *predict* and *prediction* are mentioned in fewer than 5 percent of abstracts over the past 10 years in various flagship journals in economics, political science, and sociology, and of the articles mentioning either term, only 13 percent proceed to generate actual predictions of the outcome variable (Table 1).[1,2]

[1]In most of the articles that mention the term *predict* or *prediction*, the authors use the commonplace, conceptual meaning term (e.g., "we predict that" or "our theory makes several predictions"). The actual process of making predictions of the outcome variable is virtually nonexistent in the literature cited. Note that the term *explain* or *explanation* features in only about 13 percent of abstracts, although this proportion likely does not reflect the proportion of work that is explanatory. Explanation is the default approach to empirical work, making it less relevant to explicitly mention the term in the abstract.
[2]The single sociological article making predictions in fact generated mortality forecasts (Miech et al. 2011).

[1]Nuffield College, University of Oxford, Oxford, UK
[2]Leverhulme Centre for Demographic Science, University of Oxford, Oxford, UK

Corresponding Author:
Mark D. Verhagen, Nuffield College, University of Oxford, 1 New Road, Oxford OX1 1NF, UK.
Email: mark.verhagen@nuffield.ox.ac.uk

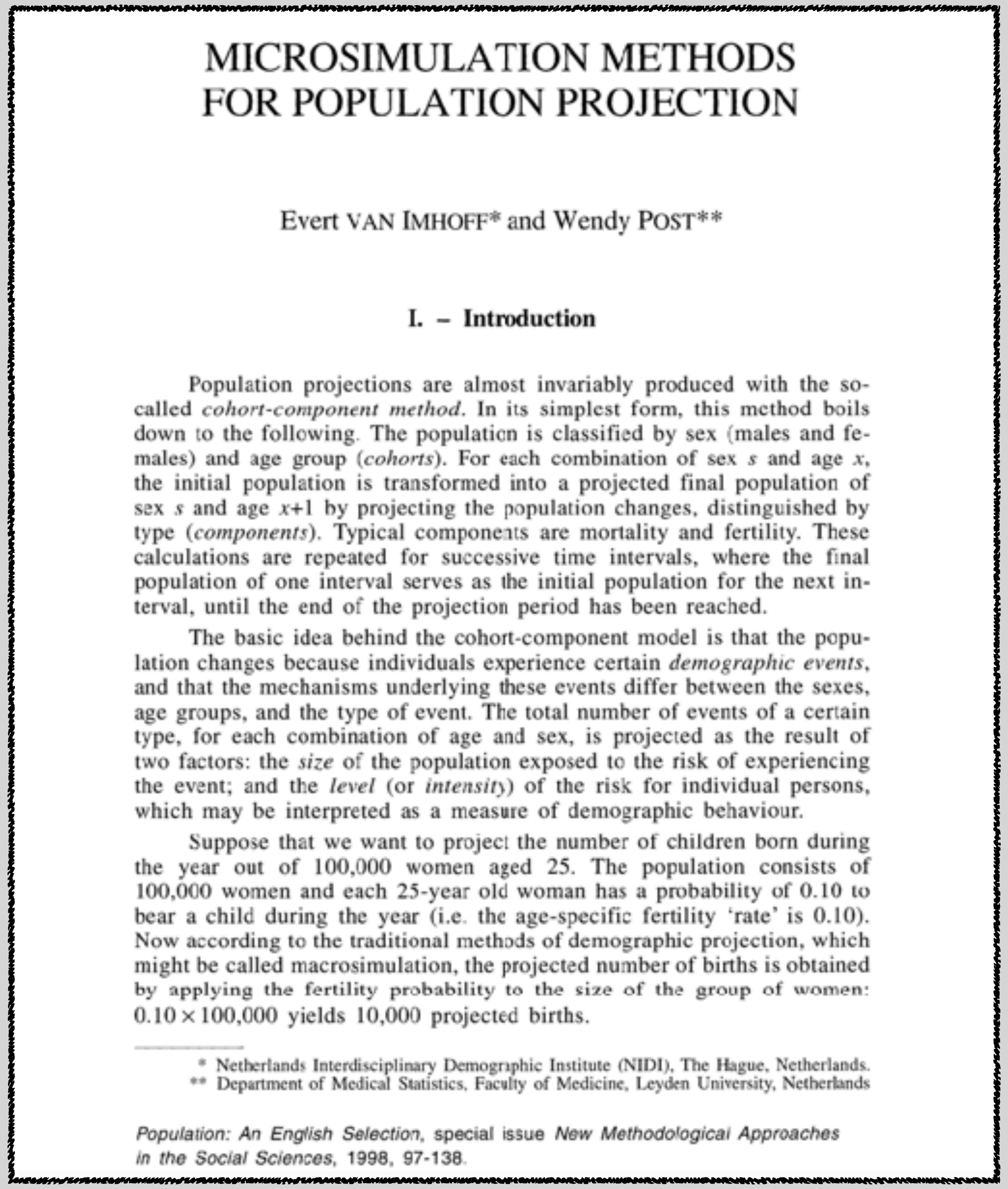
The Proposal

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

microsimulation can advance traditional statistical modelling

The Proposal

microsimulation can advance traditional statistical modelling



MICROSIMULATION METHODS
FOR POPULATION PROJECTION

Evert VAN IMHOFF* and Wendy POST**

I. – **Introduction**

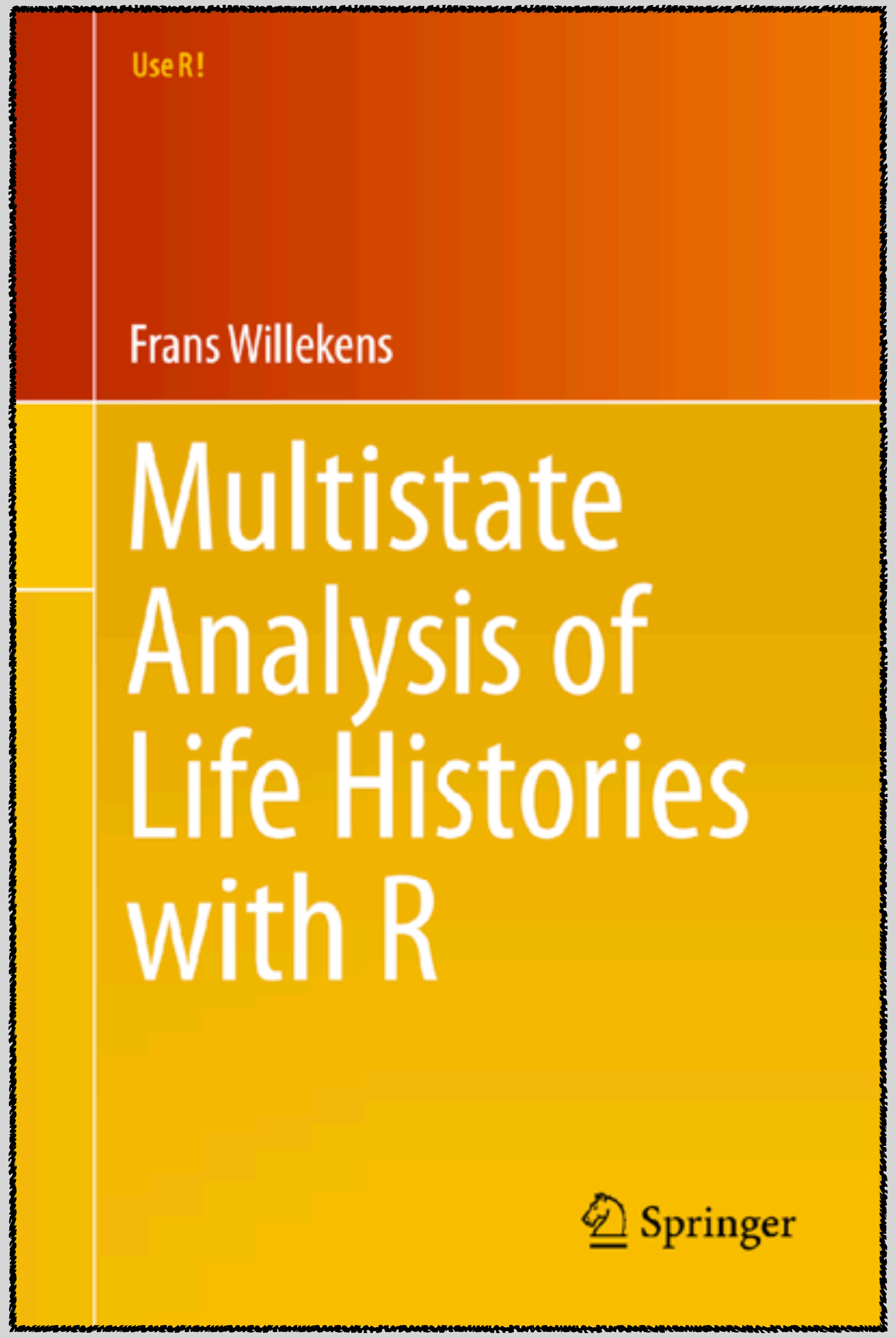
Population projections are almost invariably produced with the so-called *cohort-component method*. In its simplest form, this method boils down to the following. The population is classified by sex (males and females) and age group (*cohorts*). For each combination of sex s and age x, the initial population is transformed into a projected final population of sex s and age x+1 by projecting the population changes, distinguished by type (*components*). Typical components are mortality and fertility. These calculations are repeated for successive time intervals, where the final population of one interval serves as the initial population for the next interval, until the end of the projection period has been reached.

The basic idea behind the cohort-component model is that the population changes because individuals experience certain *demographic events*, and that the mechanisms underlying these events differ between the sexes, age groups, and the type of event. The total number of events of a certain type, for each combination of age and sex, is projected as the result of two factors: the *size* of the population exposed to the risk of experiencing the event; and the *level* (or *intensity*) of the risk for individual persons, which may be interpreted as a measure of demographic behaviour.

Suppose that we want to project the number of children born during the year out of 100,000 women aged 25. The population consists of 100,000 women and each 25-year old woman has a probability of 0.10 to bear a child during the year (i.e. the age-specific fertility 'rate' is 0.10). Now according to the traditional methods of demographic projection, which might be called macrosimulation, the projected number of births is obtained by applying the fertility probability to the size of the group of women: $0.10 \times 100{,}000$ yields 10,000 projected births.

* Netherlands Interdisciplinary Demographic Institute (NIDI), The Hague, Netherlands.
** Department of Medical Statistics, Faculty of Medicine, Leyden University, Netherlands

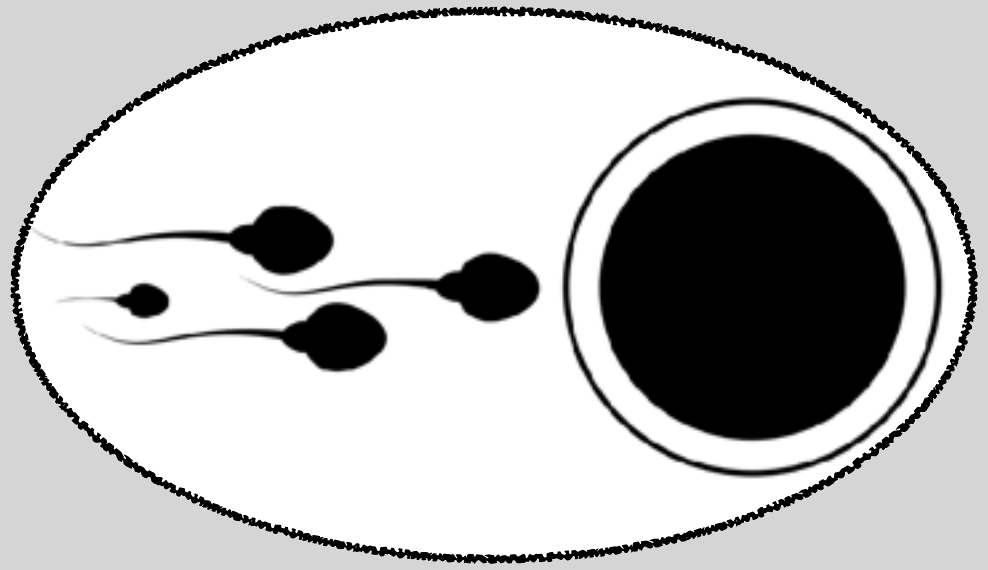
Population: An English Selection, special issue *New Methodological Approaches in the Social Sciences*, 1998, 97-138.



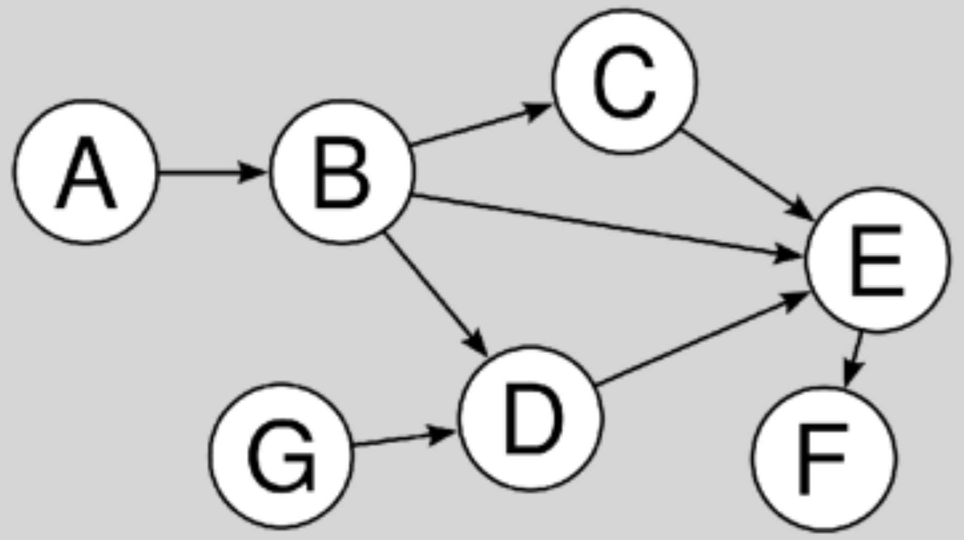
The Proposal

microsimulation can
advance traditional
statistical modelling

microsimulation can:



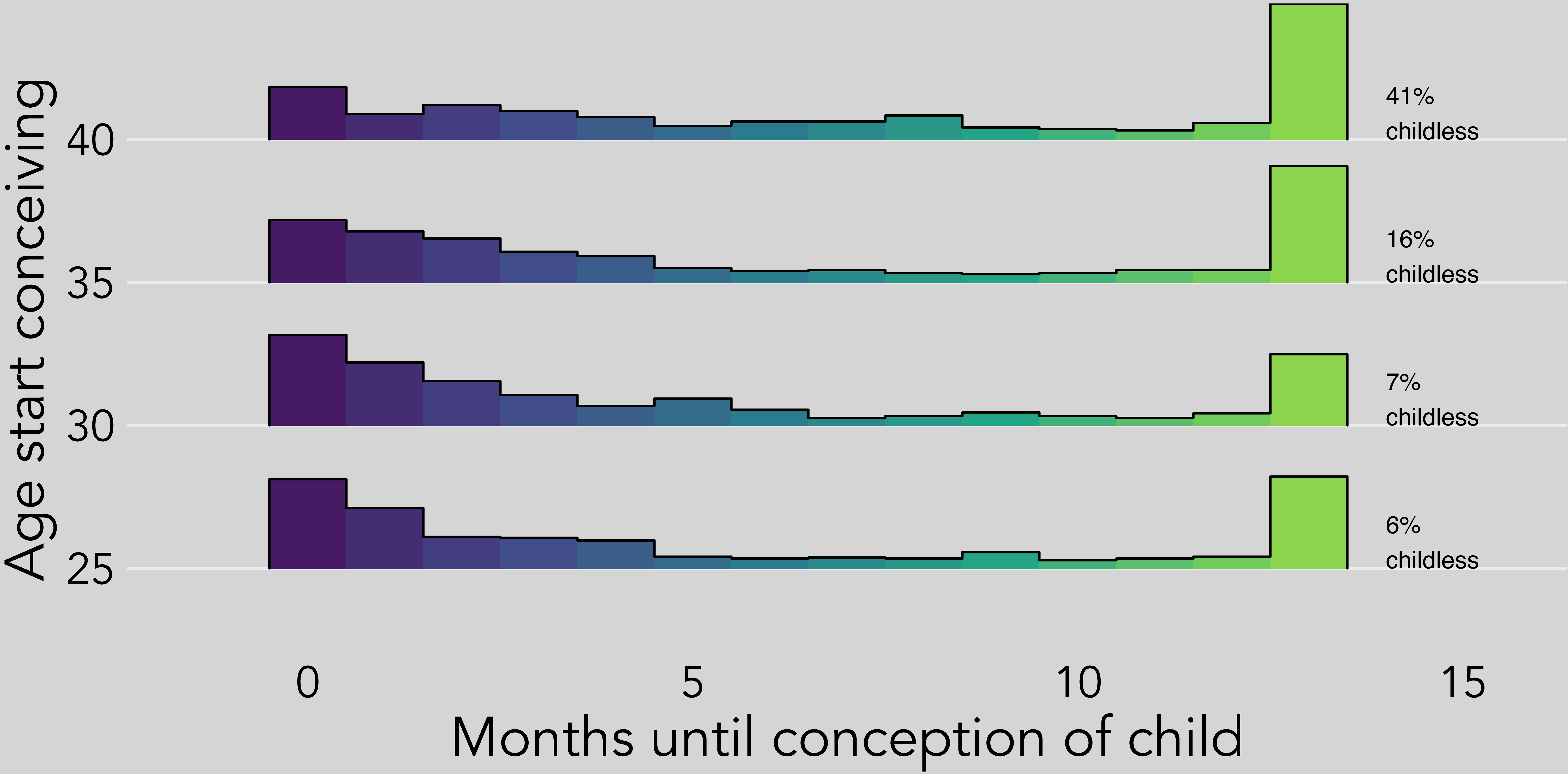
include
biological
information

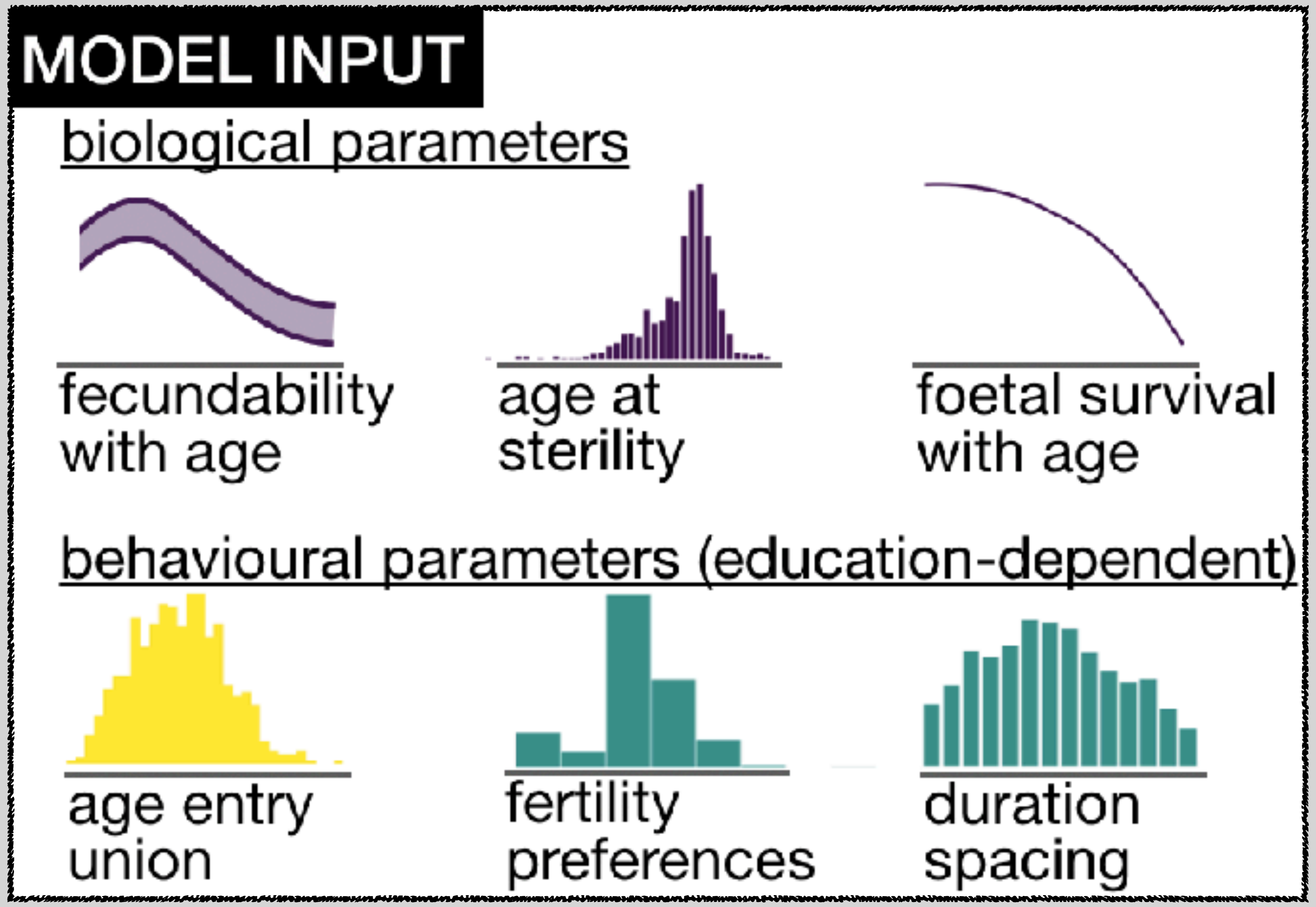


test (causal)
mechanisms

quantify
unpredictability


MODEL INPUT
biological parameters
fecundability with age
age at sterility
foetal survival with age


Age start conceiving
40
35
30
25
41% childless
16% childless
7% childless
6% childless
0
5
10
15
Months until conception of child



determines whether and when people would like to conceive

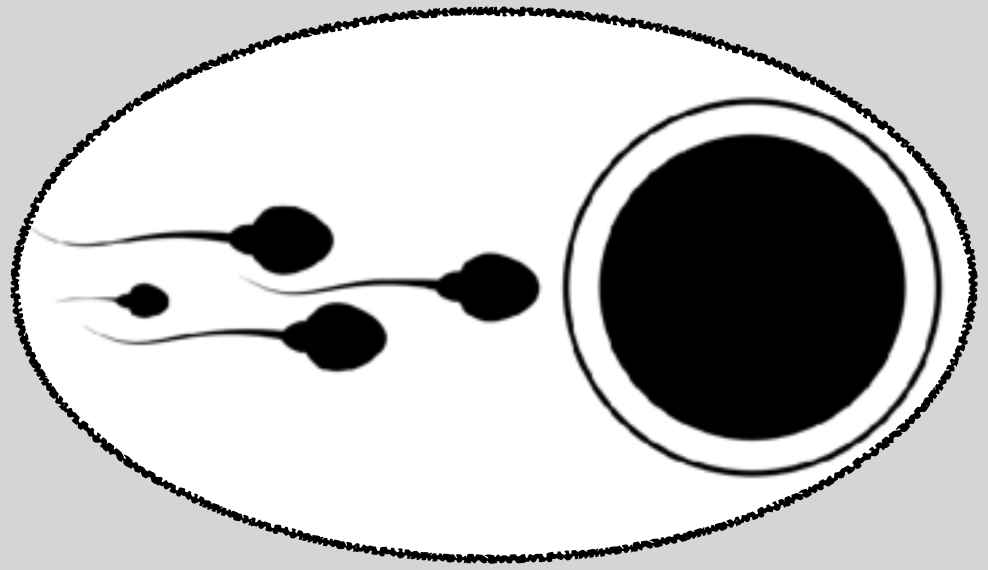
determines whether and when people conceive


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
stochasticity
stochasticity
1
22 union
27 start conceiving
29 birth
34 start conceiving
38 miscarriage
43 sterility
Randomly determined traits individual 10000
in union =26 | spac. =1 | pref. =1 | fecund. =0.1 | steril. =45 | edu. =low
stochasticity
10000
26 union
27 start conceiving
31 birth
no further children preferred
due to:
partner status
preferences
stochasticity
% childless
low
high
education
frequency
age first birth
frequency
children

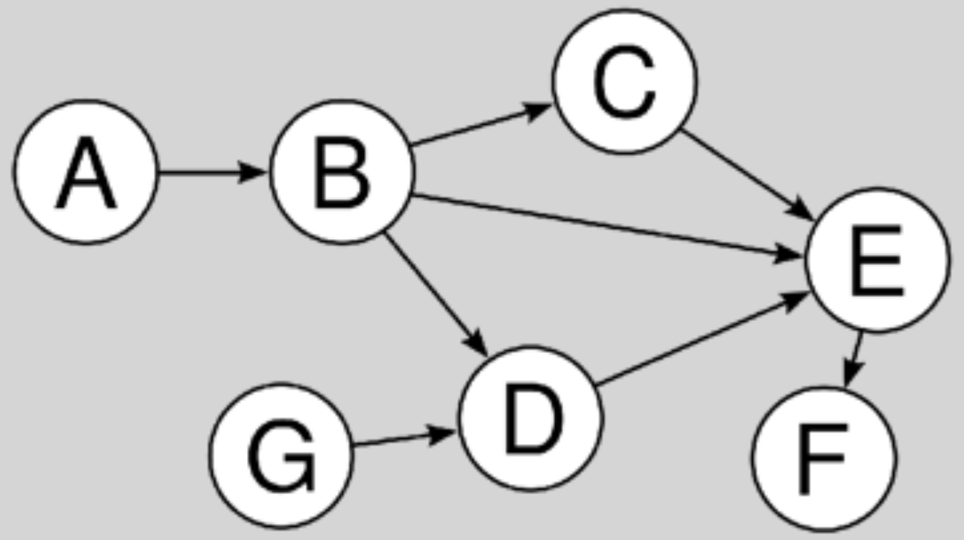
The Proposal

microsimulation can
advance traditional
statistical modelling

microsimulation can:



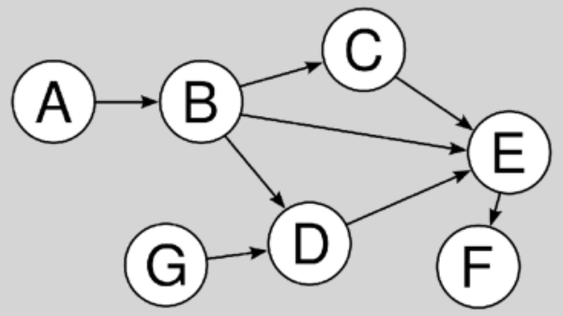
include
biological
information

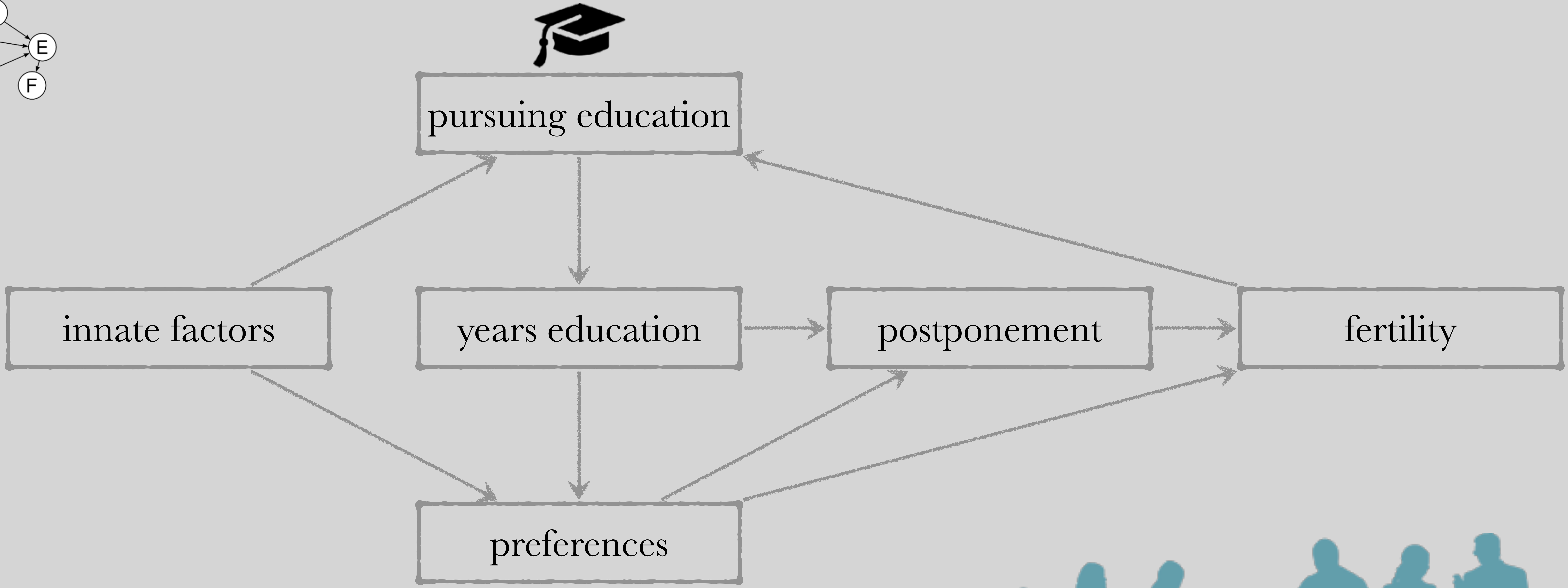


test (causal)
mechanisms

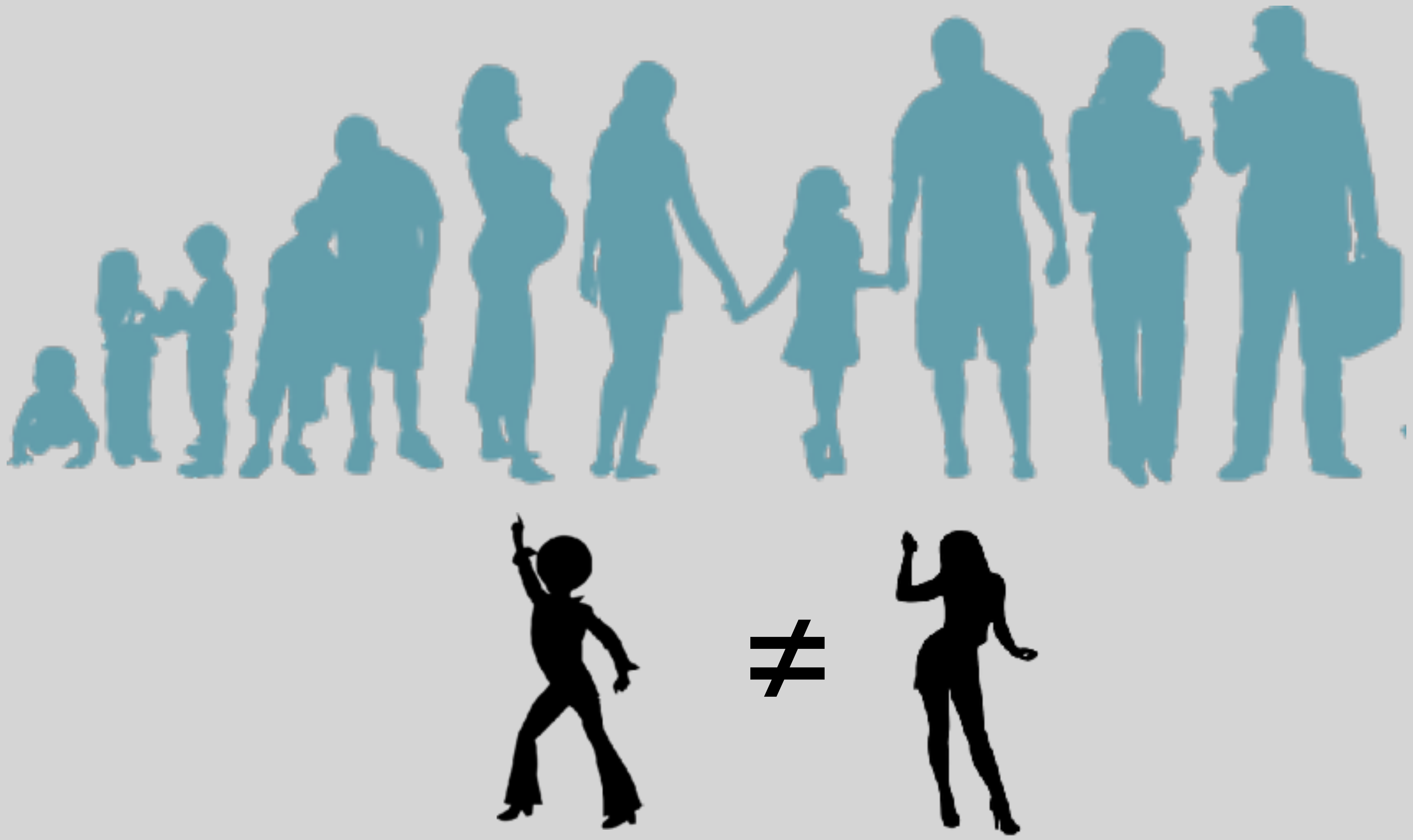
quantify
unpredictability


A
B
C
D
E
F
G
pursuing education
innate factors
years education
postponement
fertility
preferences
PNAS
Childbearing impeded education more than education impeded childbearing among Norwegian women
Joel E. Cohen[a,1], Øystein Kravdal[b], and Nico Keilman[b]
[a]Laboratory of Populations, The Rockefeller University and Columbia University, New York, NY 10065-5399; and [b]Department of Economics, University of Oslo, Blindern, 0317 Oslo, Norway
Contributed by Joel E. Cohen, June 1, 2011 (sent for review March 29, 2011)
In most societies, women at age 39 with higher levels of education have fewer children. To understand this association, we investi-
about effects of childbearing on education), probably also affect the woman's childbearing intentions. Furthermore, the resources





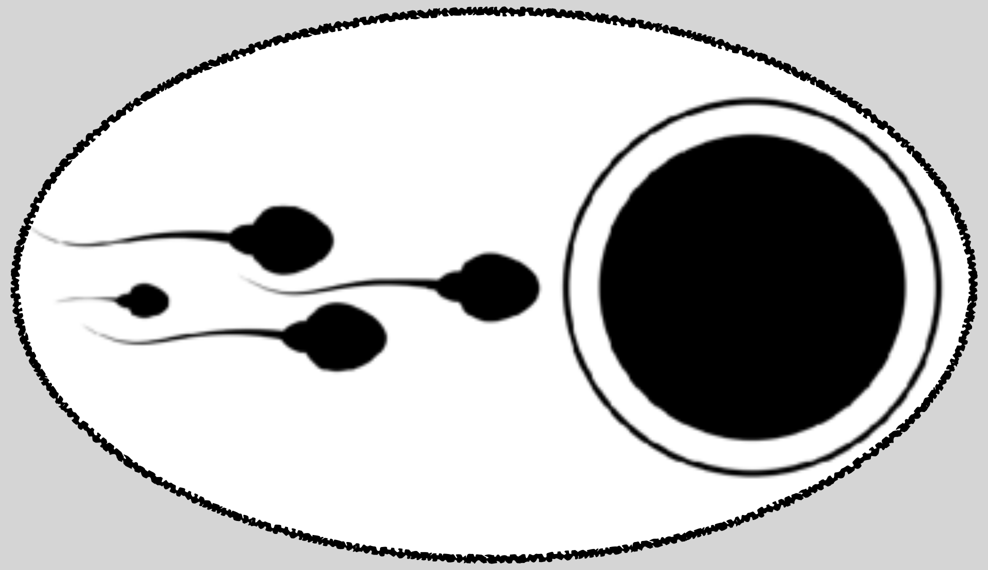
What Kind of Data Would We need to Address This Model?



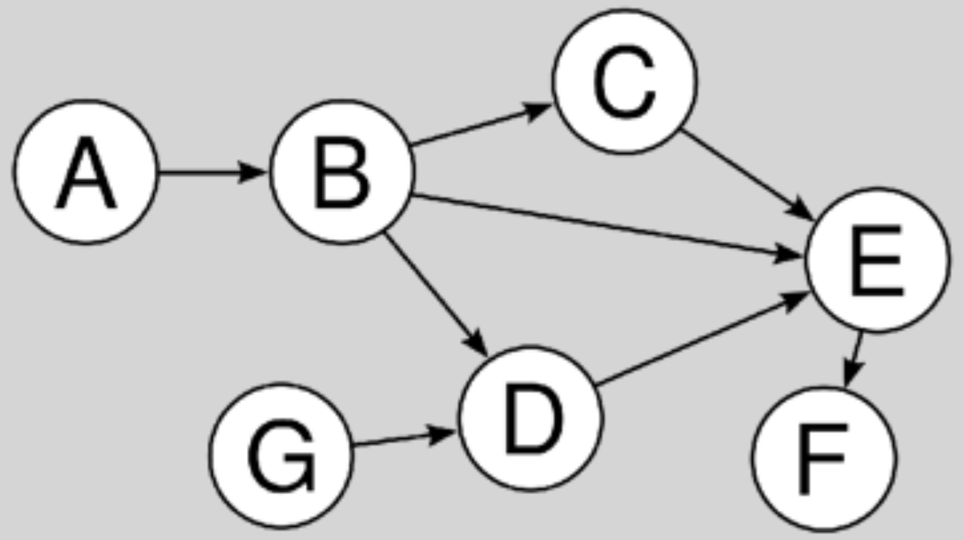
The Proposal

microsimulation can advance traditional statistical modelling

microsimulation can:



include biological information

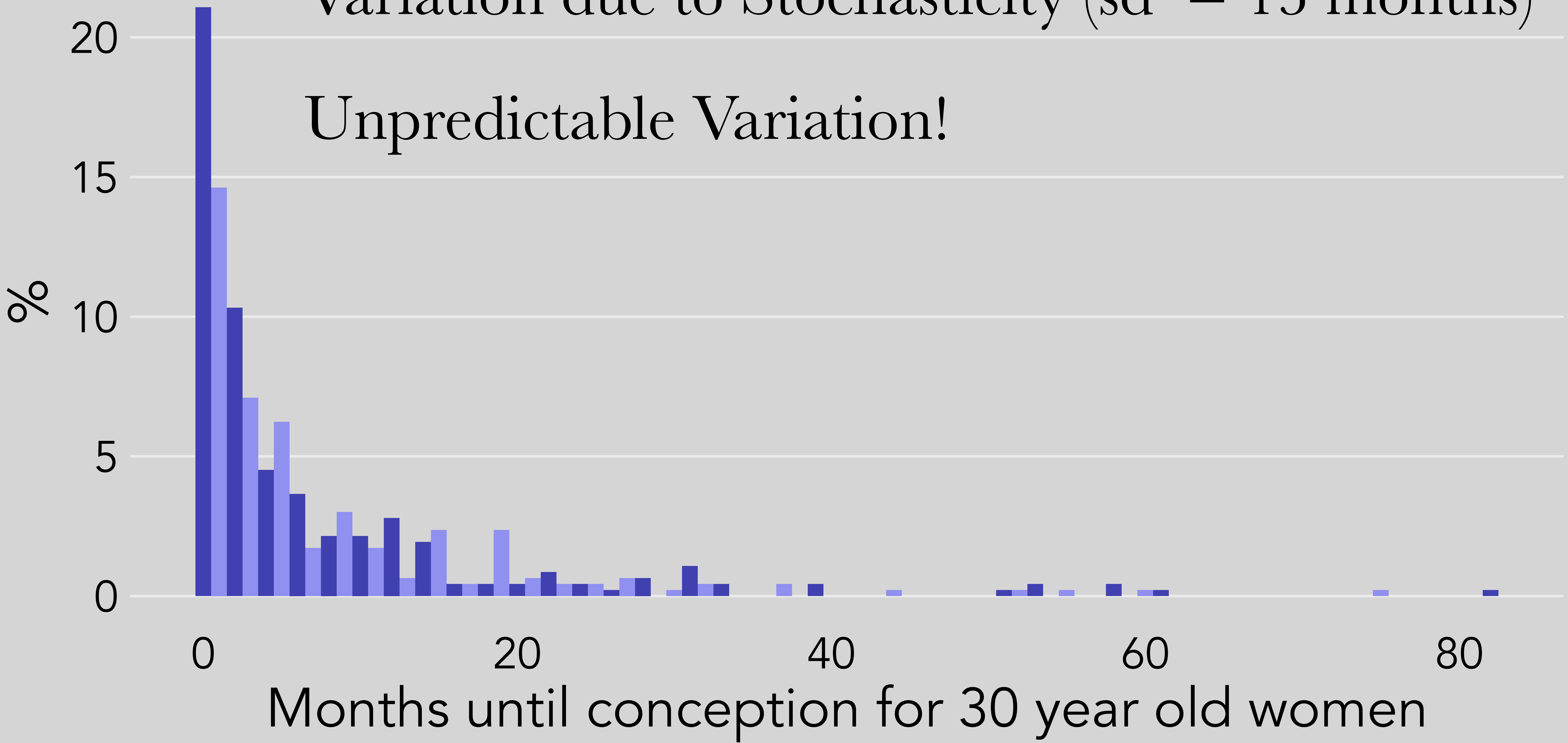


test (causal) mechanisms

quantify unpredictability

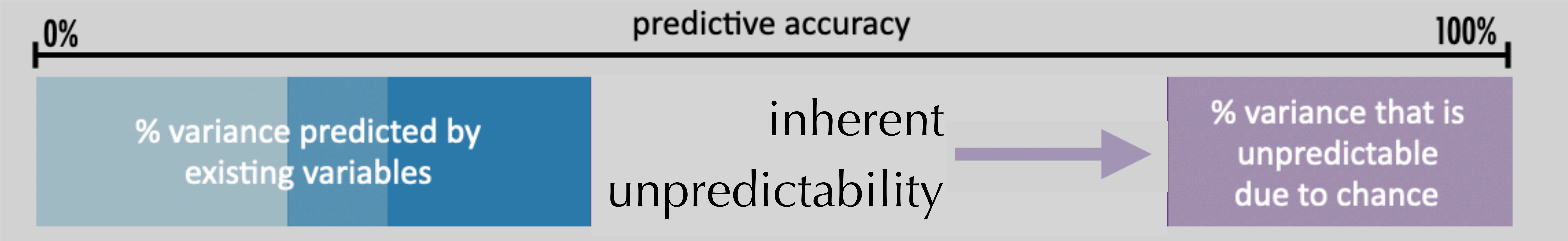
Variation due to Stochasticity (sd = 13 months)

Unpredictable Variation!

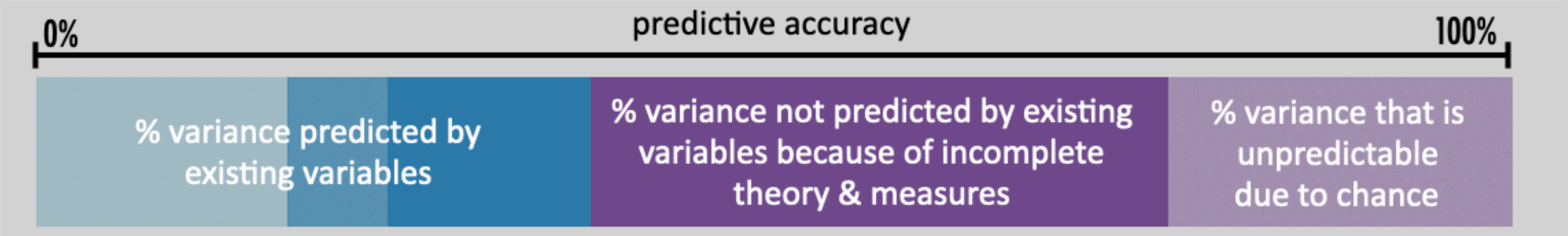


MODEL INPUT
biological parameters
fecundability with age
age at sterility
foetal survival with age
behavioural parameters (education-dependent)
age entry union
fertility preferences
duration spacing
MODEL RUN
Randomly determined traits individual 1
in union =22 | spac. =5 | pref. =2 | fecund. =0.3 | steril. =43 | edu. =high
stochasticity
stochasticity
1
22 union
27 start conceiving
29 birth
34 start conceiving
38 miscarriage
43 sterility
Randomly determined traits individual 10000
in union =26 | spac. =1 | pref. =1 | fecund. =0.1 | steril. =45 | edu. =low
stochasticity
10000
26 union
27 start conceiving
31 birth
no further children preferred
due to:
partner status
preferences
stochasticity
% childless
low
high
education
frequency
age first birth
frequency
children

Unpredictable Variation

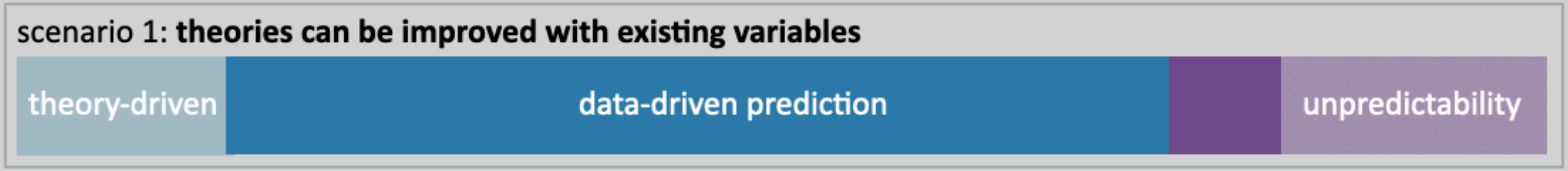


Unique Insight into State of Field

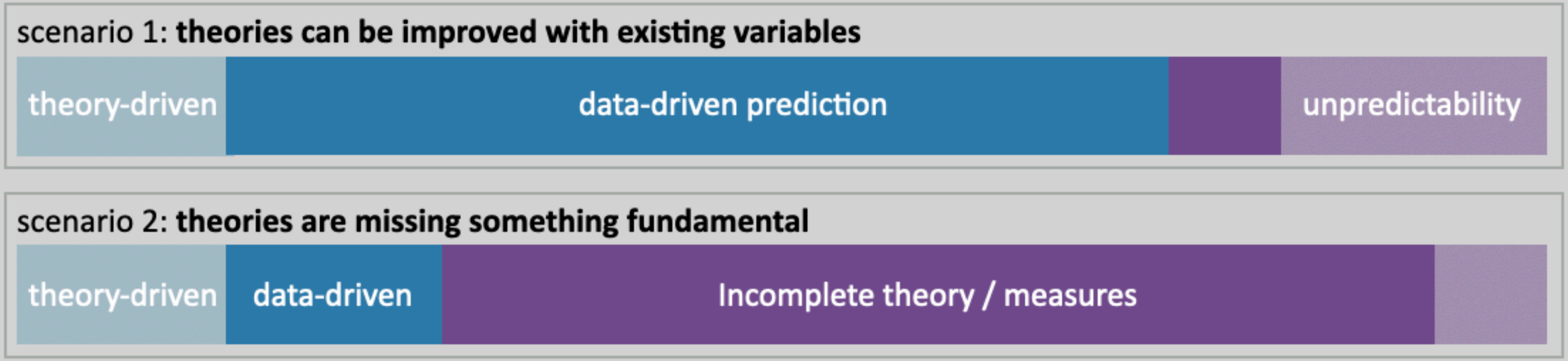


?

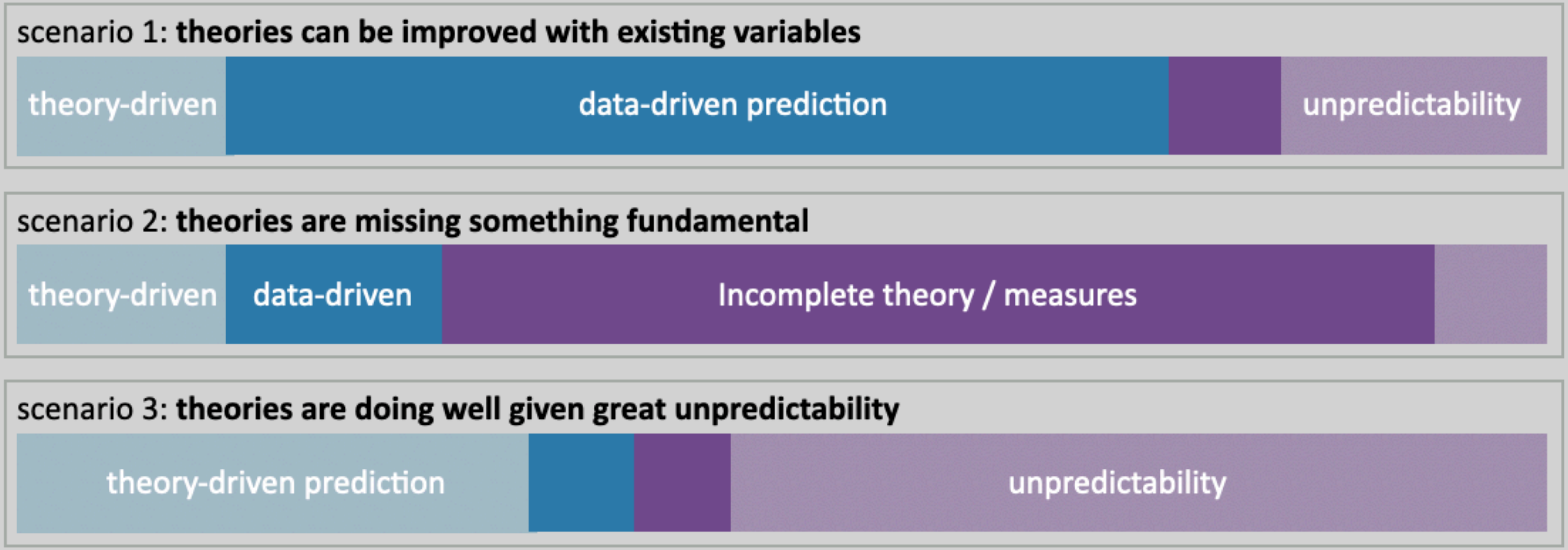
Unique Insight into State of Field



Unique Insight into State of Field



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

This mess we're in?

Or how simulation and prediction will advance fertility research

gert stulp — gertstulp.com — g.stulp@rug.nl