PreFer
Predicting Fertility data challenge

How predictable are fertility outcomes?

Introducing the PreFer data challenge and its potential for fertility research

gert stulp - university of groningen

Elizaveta Sivak

NWO

Understanding fertility outcomes by quantifing the (un)predictable

ODISSEI

Open Data Infrastructure for
Social Science and Economic Innovations

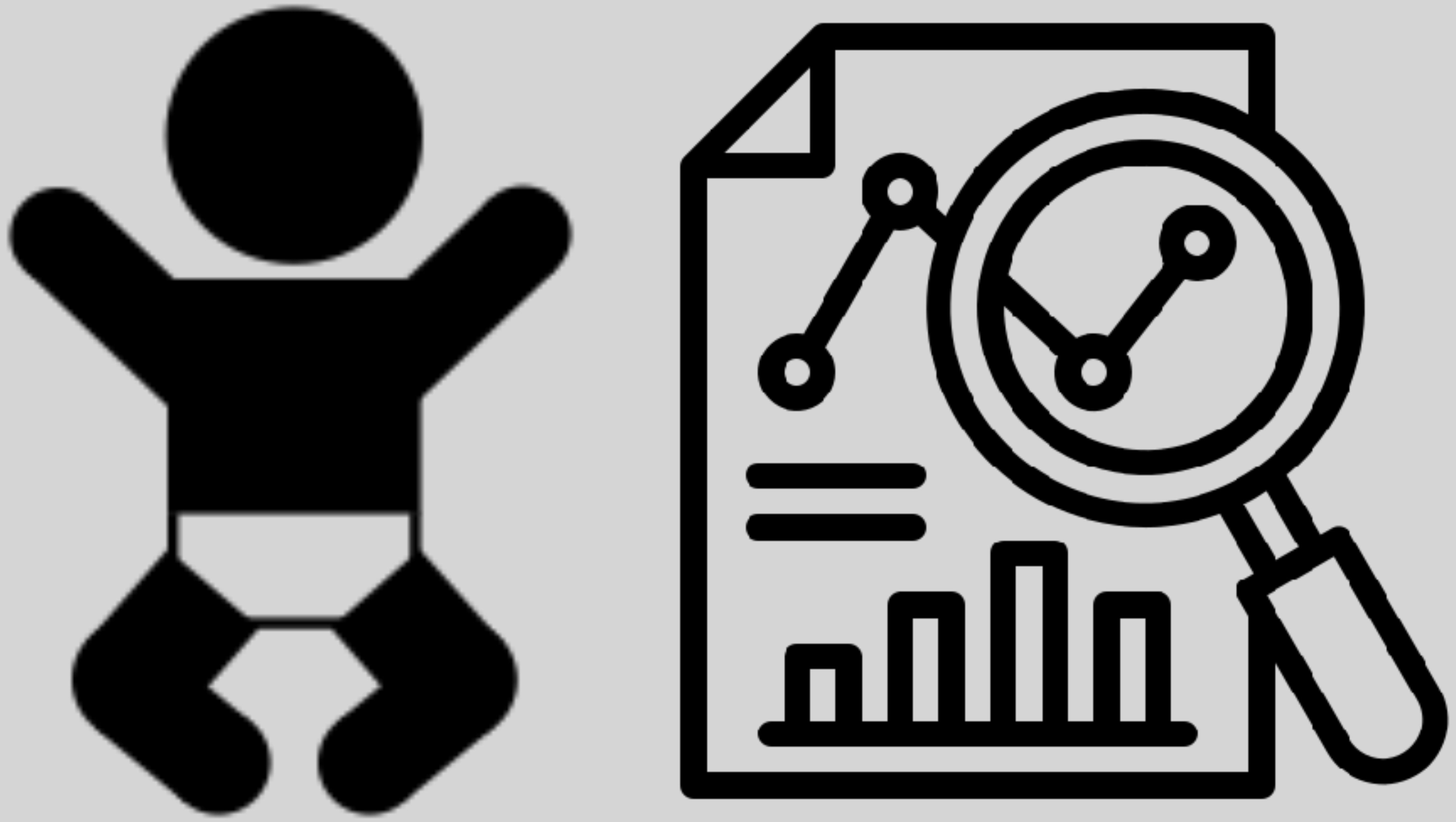
university of groningen

CENTERDATA research institute

Eyra

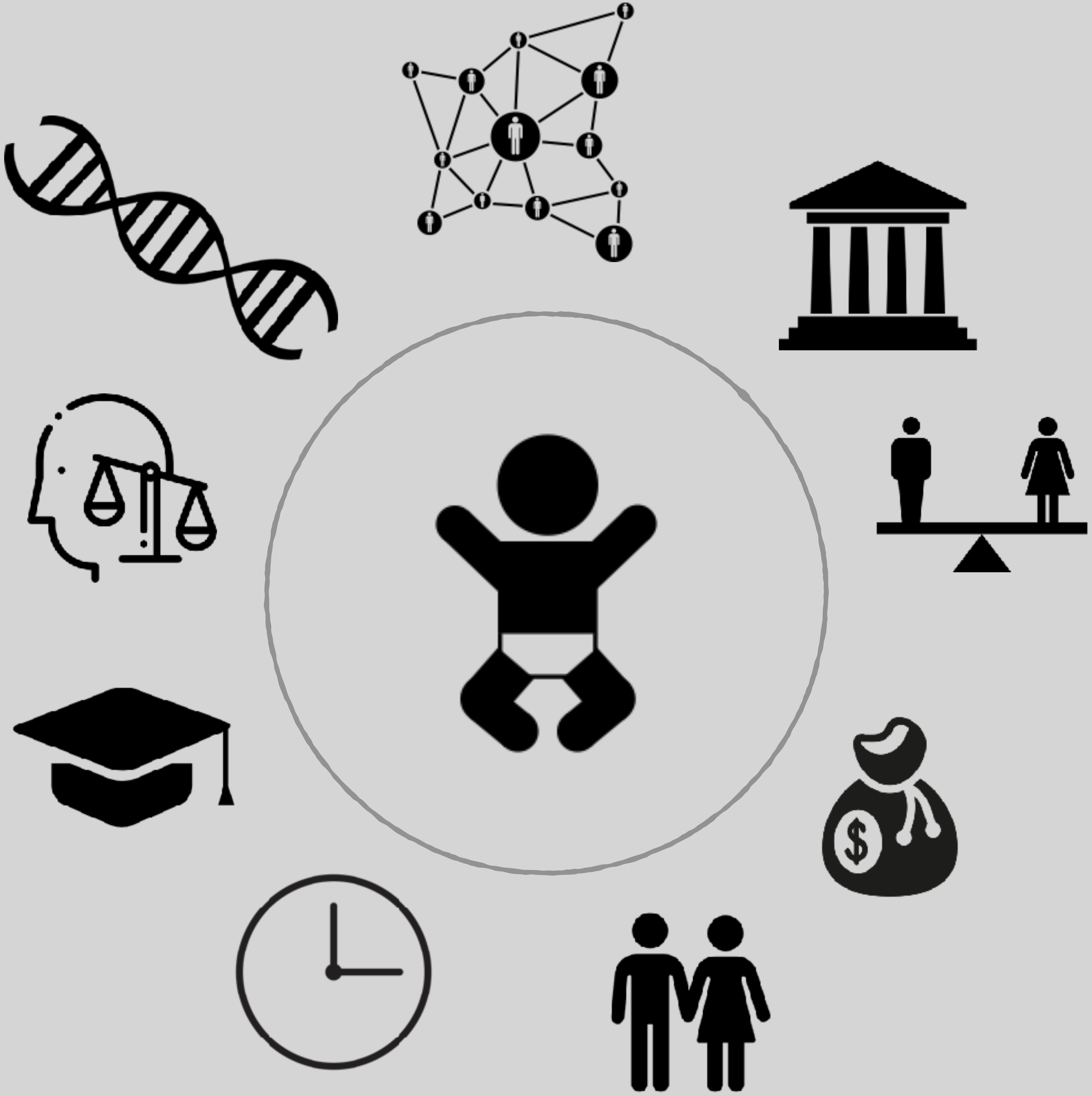
the dreamteam

Tom Emery
Javier Garcia-Bernardo
Seyit Höcük
Kasia Karpinska
Angelica Maineri
Adriënne Mendrik
Joris Mulder
Malvina Nissim
Paulina Pankowska



PreFer

Predicting Fertility data challenge



variables
explain
little

Fewer births because of study and flexwork?

“total effect on fertility … rather small

variables
explain
little

Fewer births because of study and flexwork?

“total effect on fertility … rather small

surprising
patterns

variables
explain
little

Fewer births because of study and flexwork?

“total effect on fertility … rather small

surprising
patterns

incomparable
results

variables explain little

Fewer births because of study and flexwork?

“total effect on fertility … rather small

surprising patterns

incomparable results


PHILOSOPHICAL TRANSACTIONS B

Wealth, fertility and adaptive behaviour in industrial populations

Review

1. Introduction

...ieties, Part I

...cieties, Part II

THE ROYAL SOCIETY

non-replicable results


Population Review

Volume 60, Number 2, 2021

Explaining the Associations of Education and Occupation with Childlessness: The Role of Desires and Expectations to Remain Childless

Abstract

Keywords

From Some to None? Fertility Expe... of Permanently Childless Women

Abstract

Childless Expectations and Childle... Life Course

Introduction

Replication Crisis

PSYCHOLOGY

Estimating the reproducibility of psychological science

Open Science Collaboration*

ROYAL SOCIETY OPEN SCIENCE

rsos.royalsocietypublishing.org

Research

CrossMark
click for updates

Cite this article: Smaldino PE, McElreath R. 2016 The natural selection of bad science.

The natural selection of bad science

Paul E. Smaldino[1] and Richard McElreath[2]

[1]Cognitive and Information Sciences, University of California, Merced, CA 95343, USA
[2]Department of Human Behavior, Ecology, and Culture, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

PES, 0000-0002-7133-5620; RME, 0000

Check for updates

aps ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

General Article

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

Psychological Science
22(11) 1359–1366

Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0956797611417632
http://pss.sagepub.com
SAGE

Joseph P. Simmons[1], Leif D. Nelson[2], and Uri Simonsohn[1]
[1]The Wharton School, University of Pennsylvania, and [2]Haas School of Business, University of California, Berkeley

Replication (crisis) in Family Sociology / Demography?

Reasons unlikely

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

Reasons not unlikely

Replication (crisis) in Family Sociology / Demography?

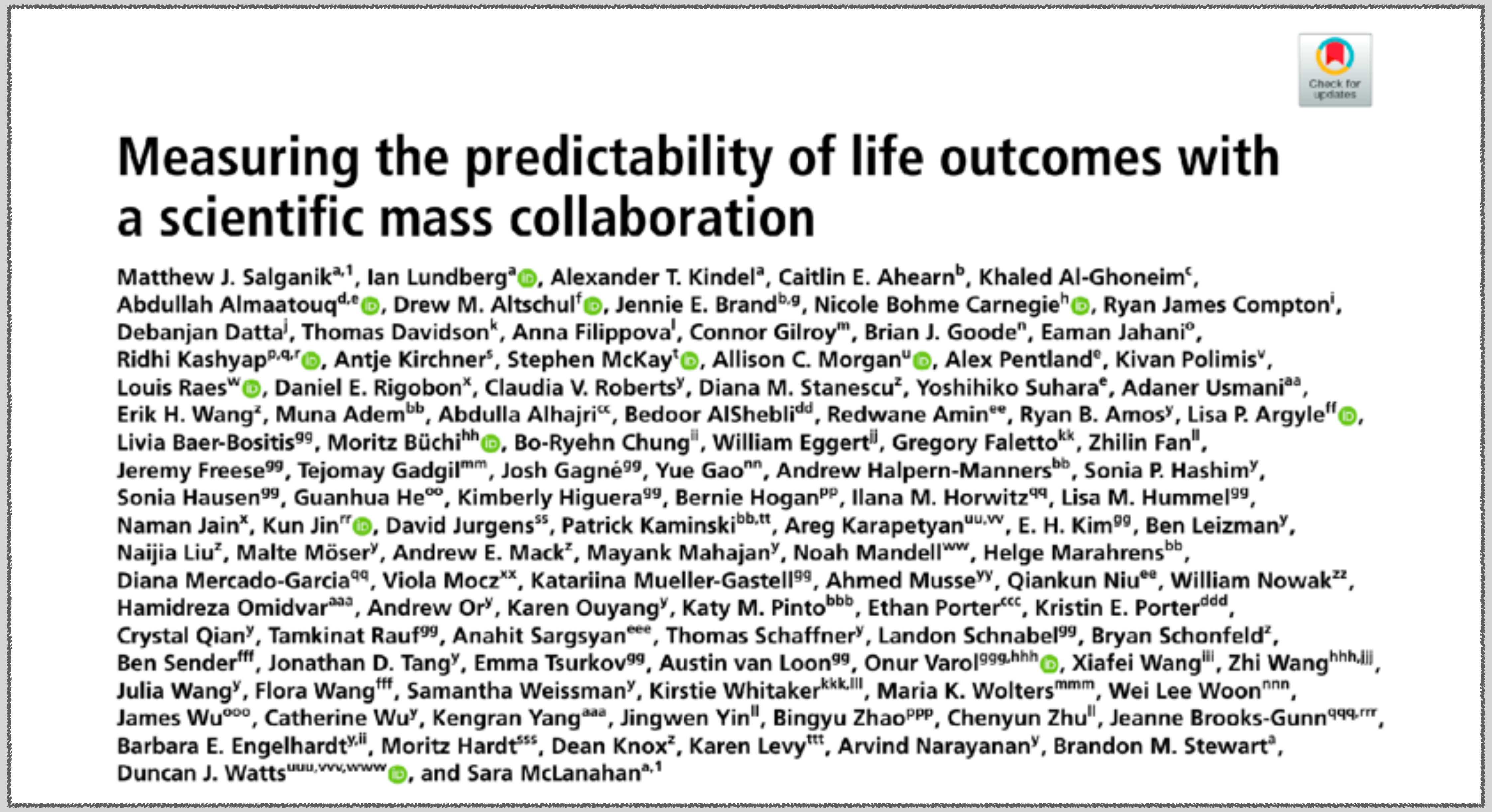
Reasons unlikely

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

Reasons not unlikely

- *Non-experimental*
- *Correlational, but little causal inference*
- *Large N, yet star gazing*
- *Controlling at will*
- *Long reign linearity*

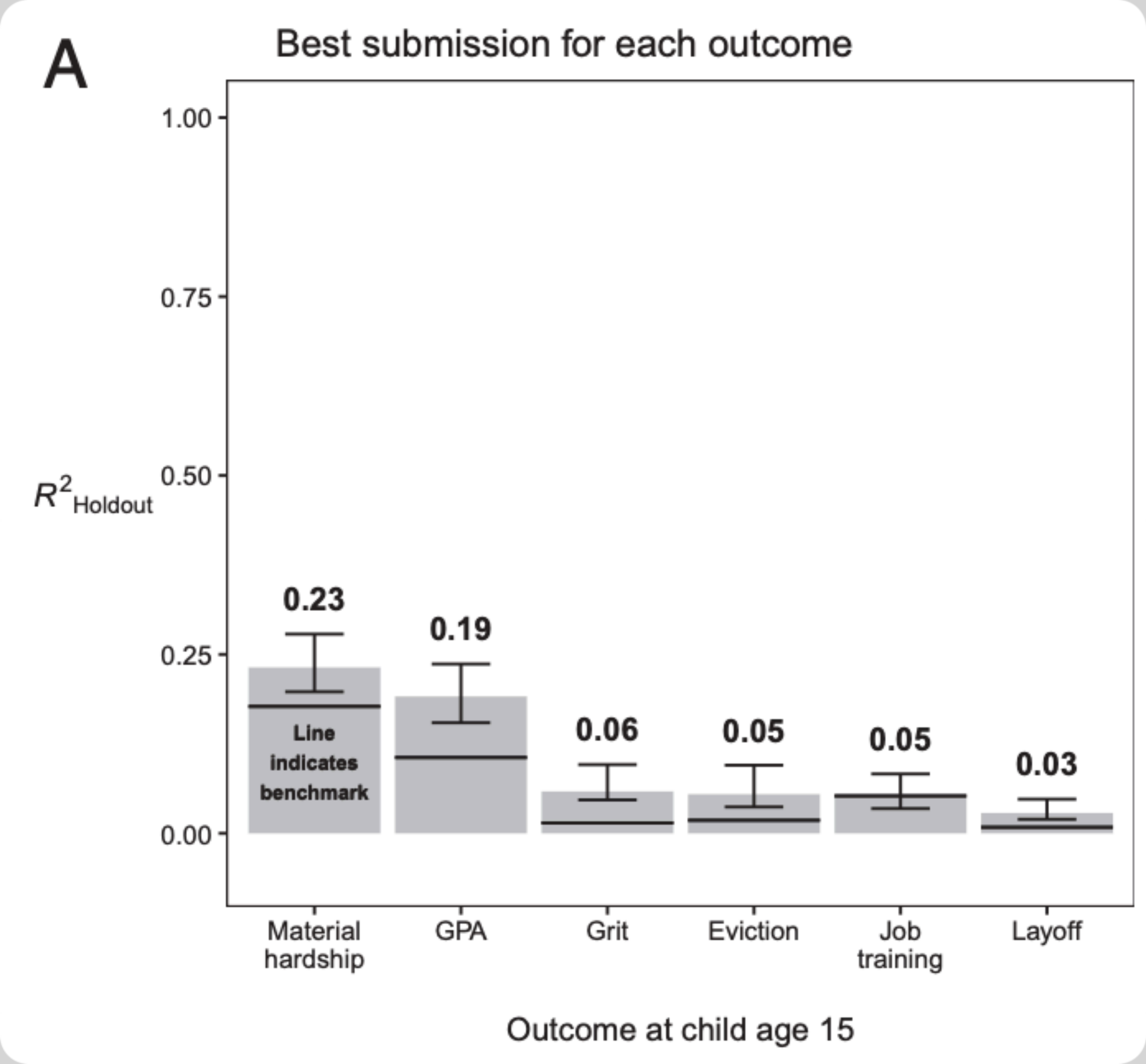
Predictability Crisis?



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?

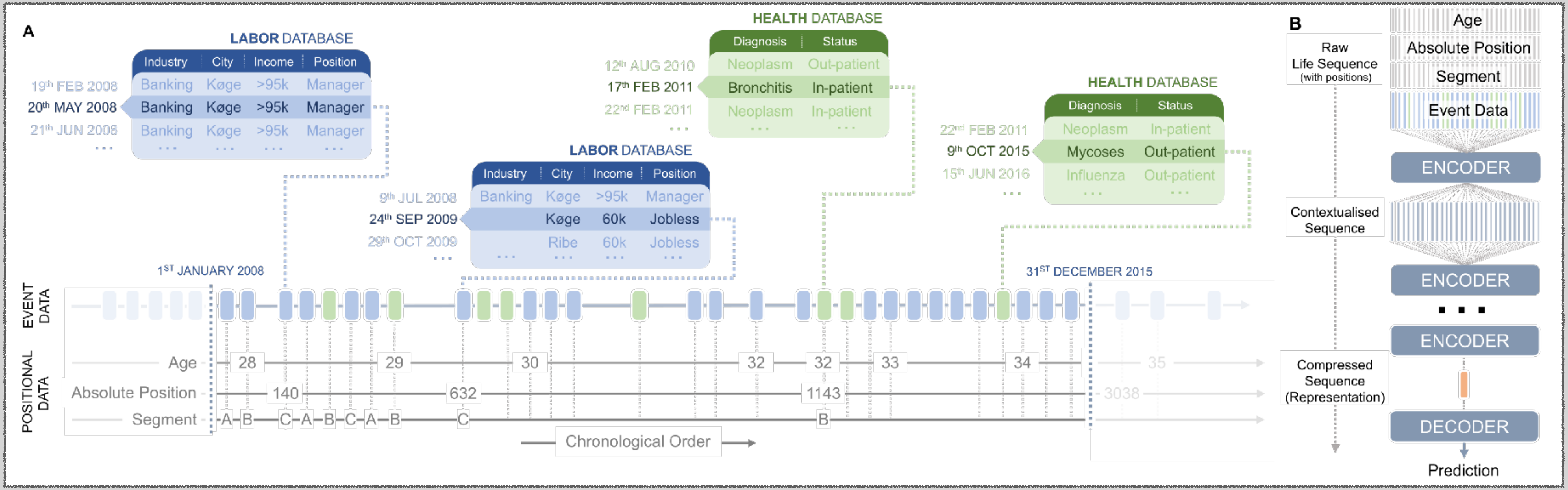
> "Social scientists studying the life course must find a way to reconcile a widespread belief that understanding has been generated by these data—as demonstrated by more than 750 published journal articles using the Fragile Families data with the fact that the very same data could not yield accurate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

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

June 6, 2023

> we show that accurate individual predictions are indeed possible



Prediction

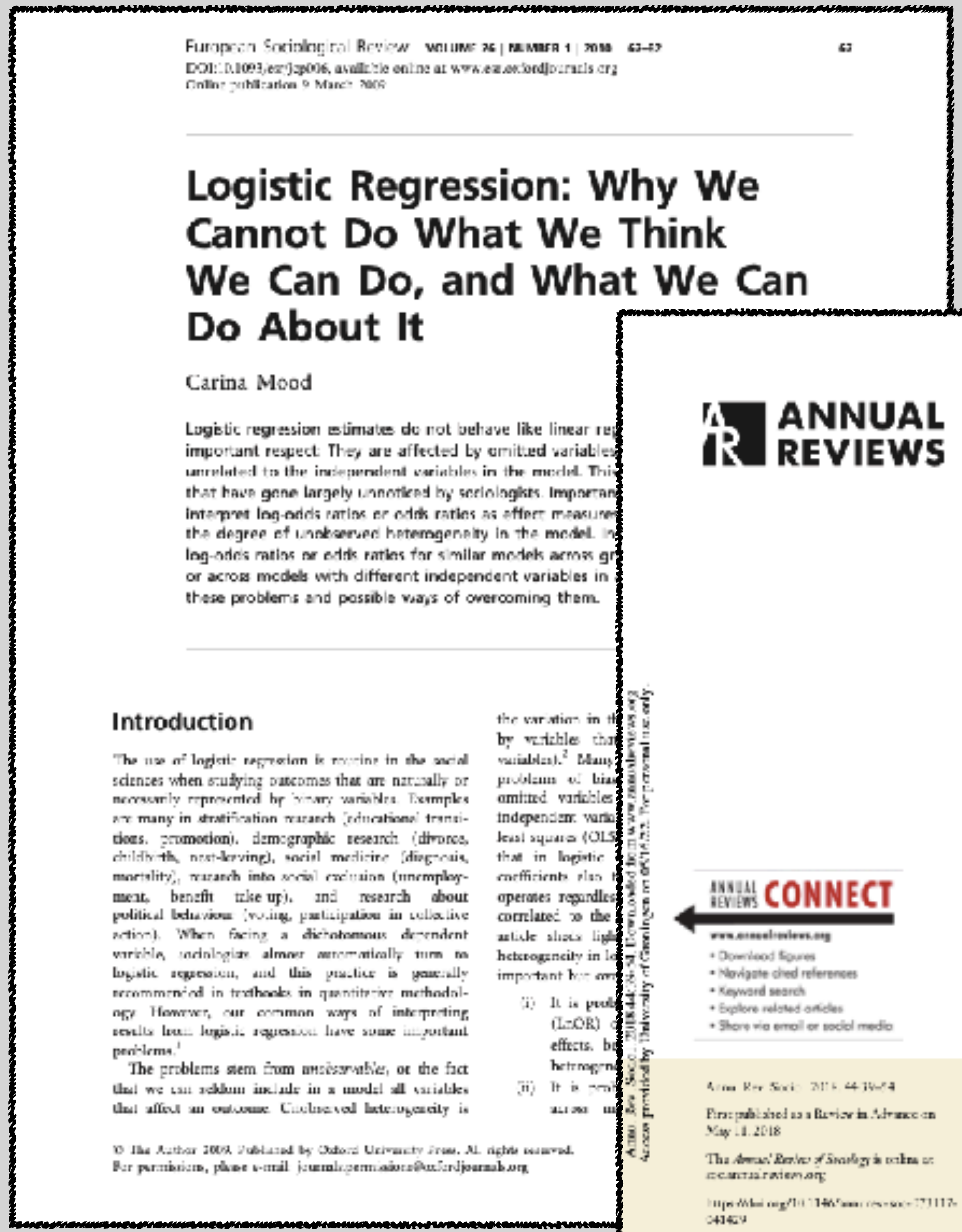
a shift towards **prediction** leads to a more reliable and useful social science

out-of-sample predictive ability:

clear measure of effect size

out-of-sample predictive ability

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



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

Carina Mood

Introduction

ANNUAL REVIEWS

Annual Review of Sociology

Interpreting and Understanding Logits, Probits, and Other Nonlinear Probability Models

Richard Breen,[1] Kristian Bernt Karlson,[2] and Anders Holm[3]

ANNUAL REVIEWS CONNECT

Keywords

logit, probit, KHB method, Y-standardization, marginal effects, linear probability model, mediation

Abstract

Methods textbooks in sociology and other social sciences routinely recommend the use of the logit or probit model when an outcome variable is binary, an ordered logit or ordered probit when it is ordinal, and a multinomial logit when it has more than two categories. But these methodological guidelines take little or no account of a body of work that, over the past 30 years, has pointed to problematic aspects of these nonlinear probability models and, particularly, to difficulties in interpreting their parameters. In this review, we draw on that literature to explain the problems, show how they manifest themselves in research, discuss the strengths and weaknesses of alternatives that have been suggested, and point to lines of further analysis.

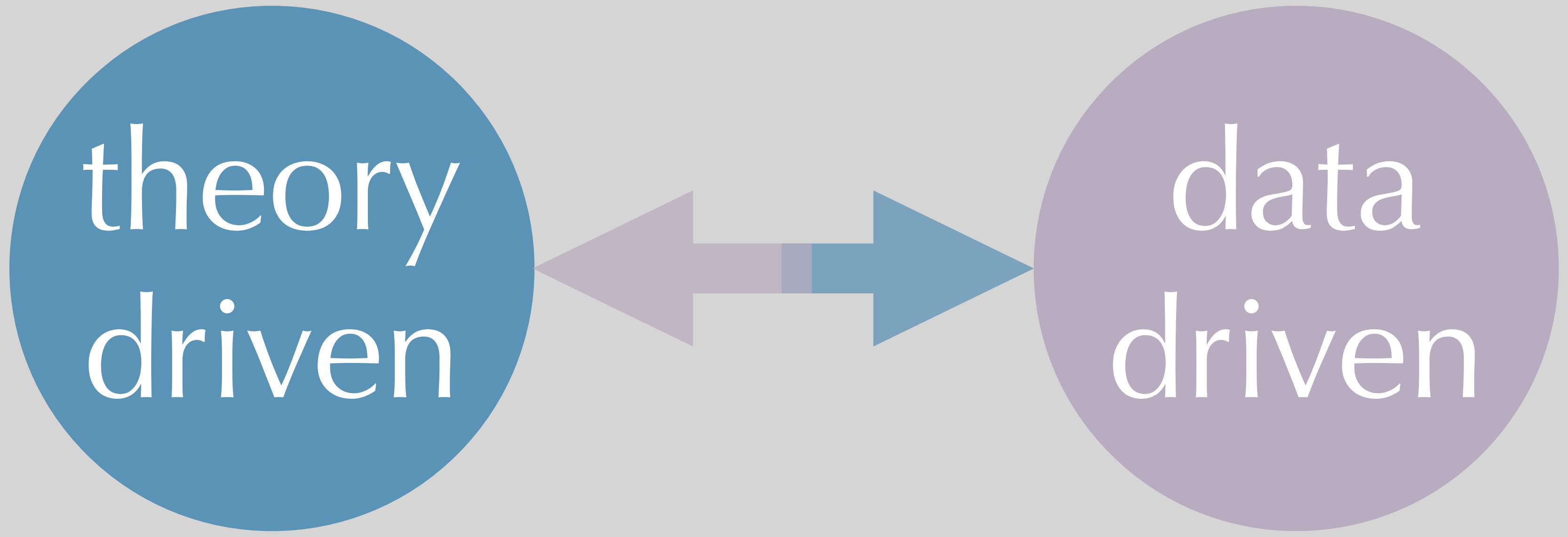
19

Prediction

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

out-of-sample predictive ability:

clear measure of effect size

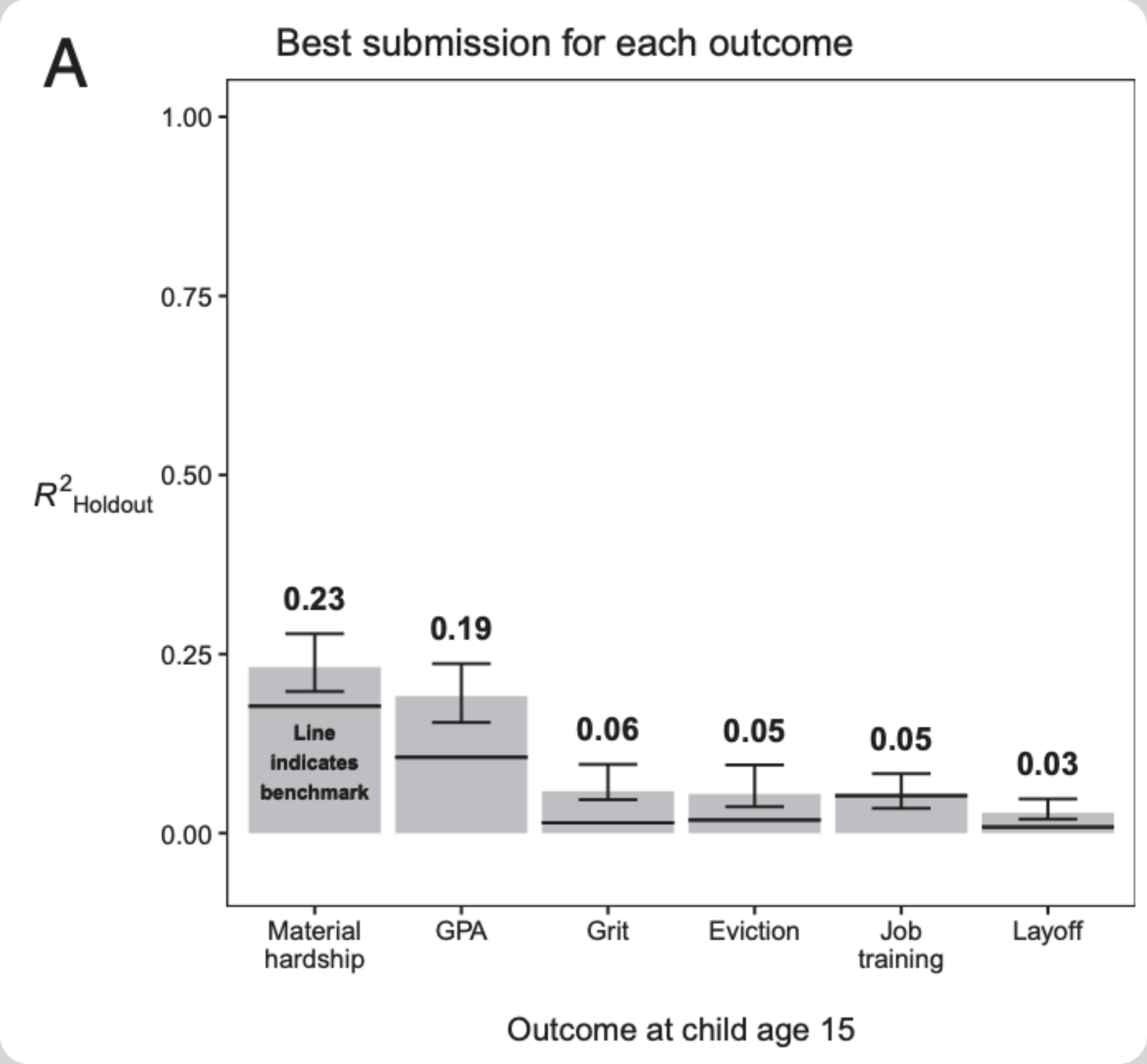


facilitates dialogue theory- and data-driven models

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]



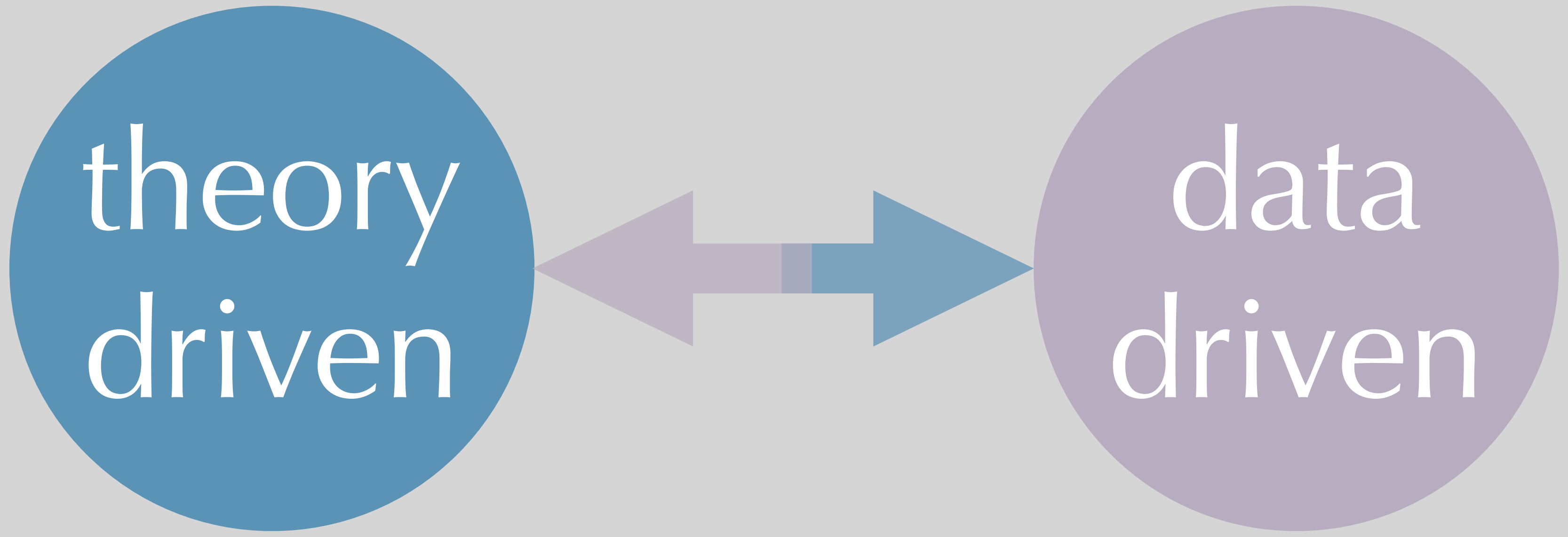
data challenge:
predicting life outcomes based on ~6000 variables by 160 teams
both theory- & data-driven

Prediction

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

Articles

Journal of Peace Research
47(4) 363–375

DOI: 10.1177/0022343309356491
jpr.sagepub.com
SAGE

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

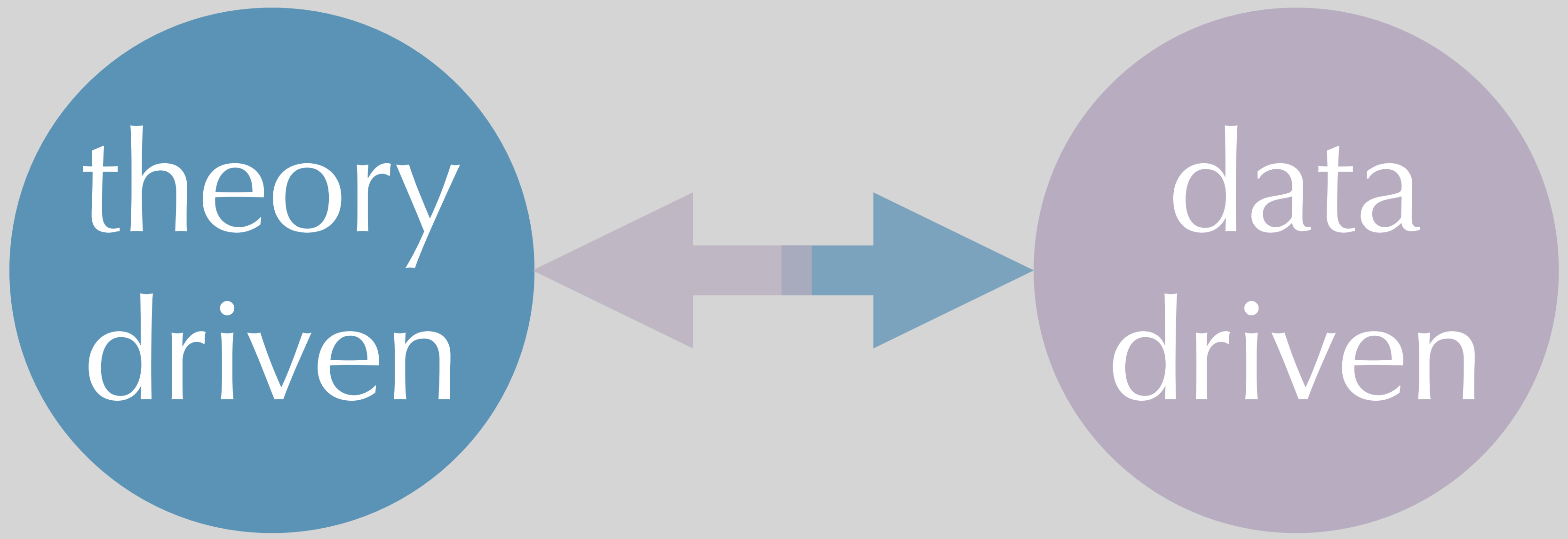
Department of Political Science, University College London

Prediction

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

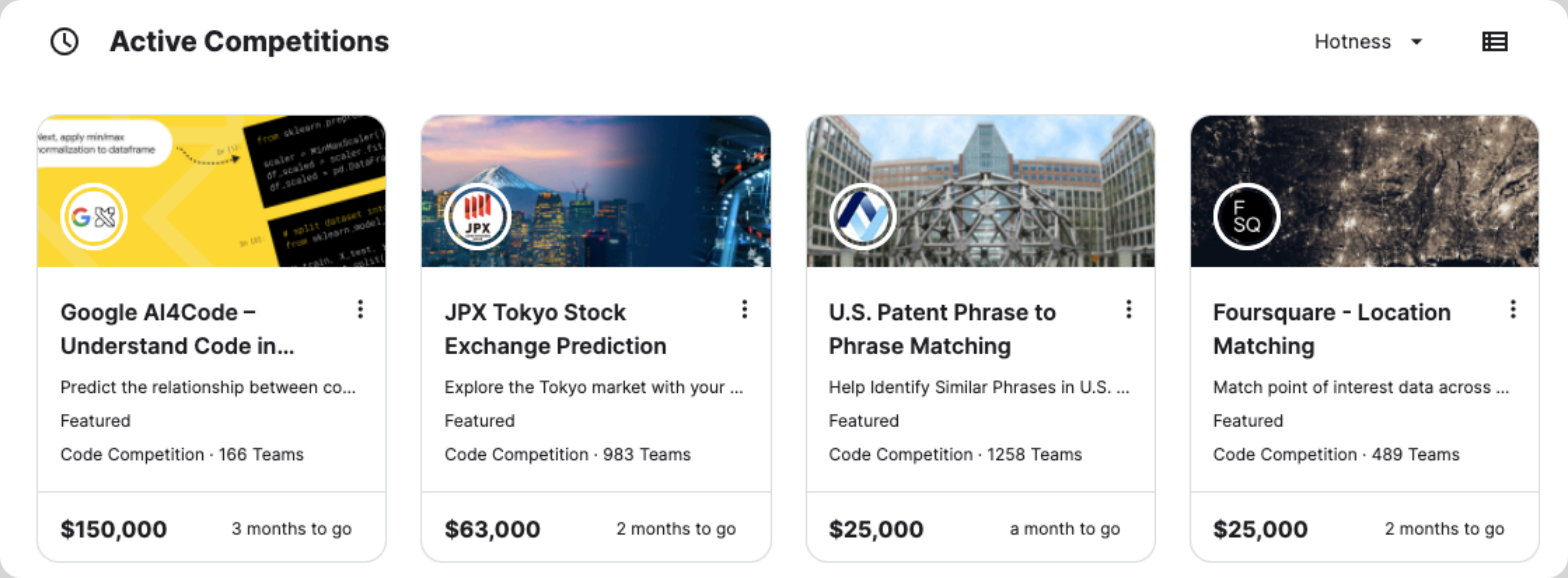
data
challenge

theory driven

data driven

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

kaggle



“secret sauce of data science

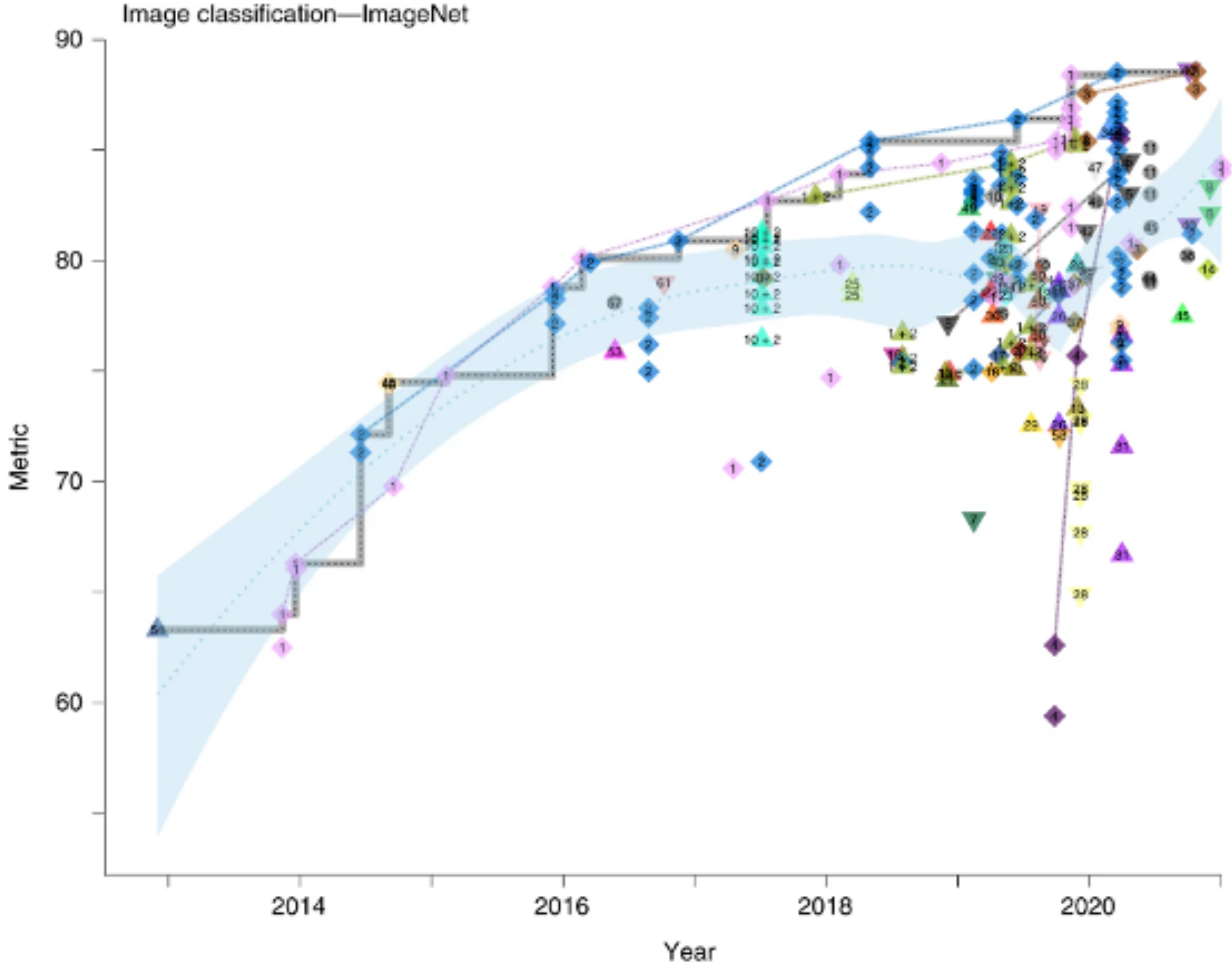
Donoho, 2015

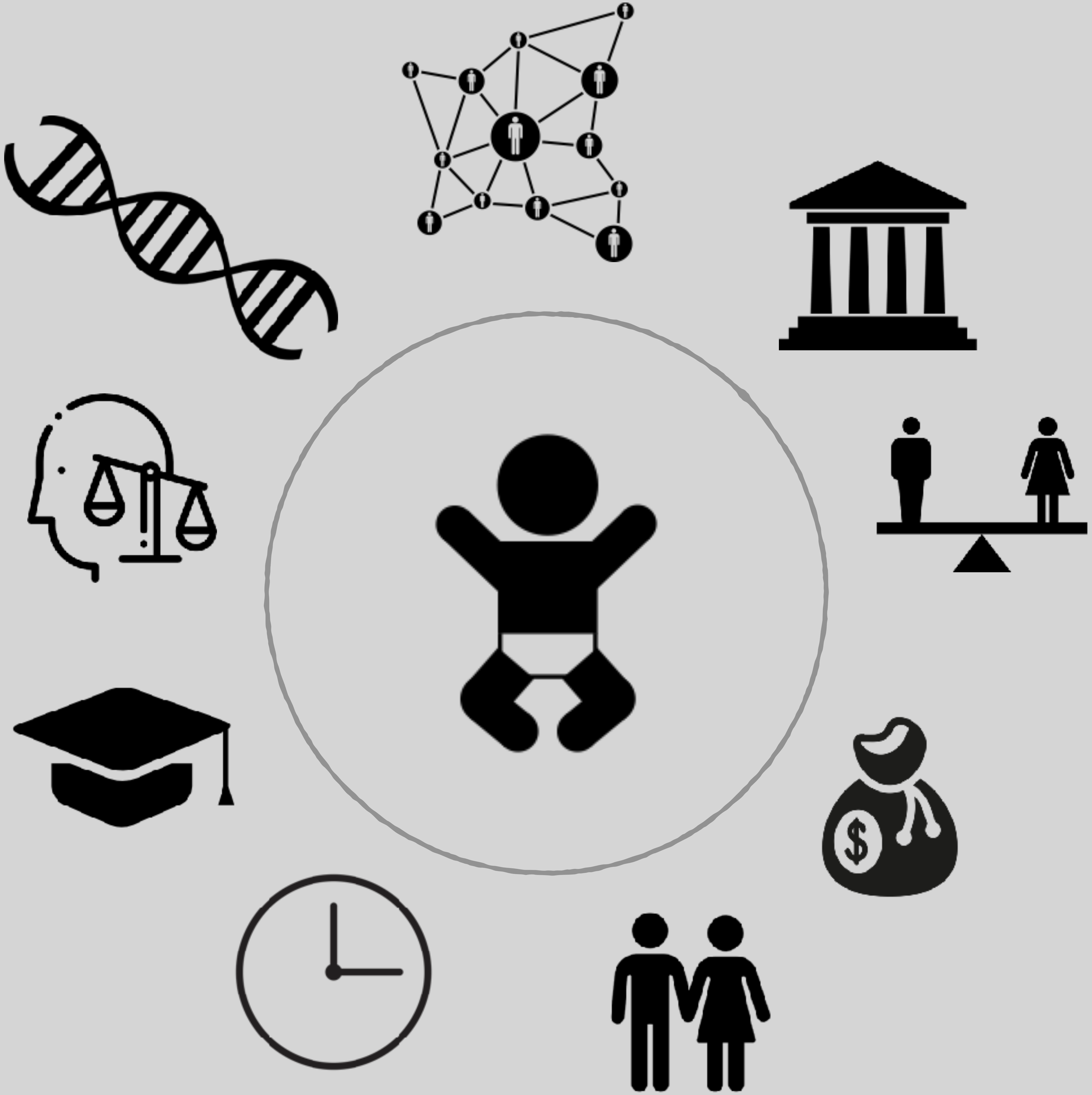
Prediction Benchmarks

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

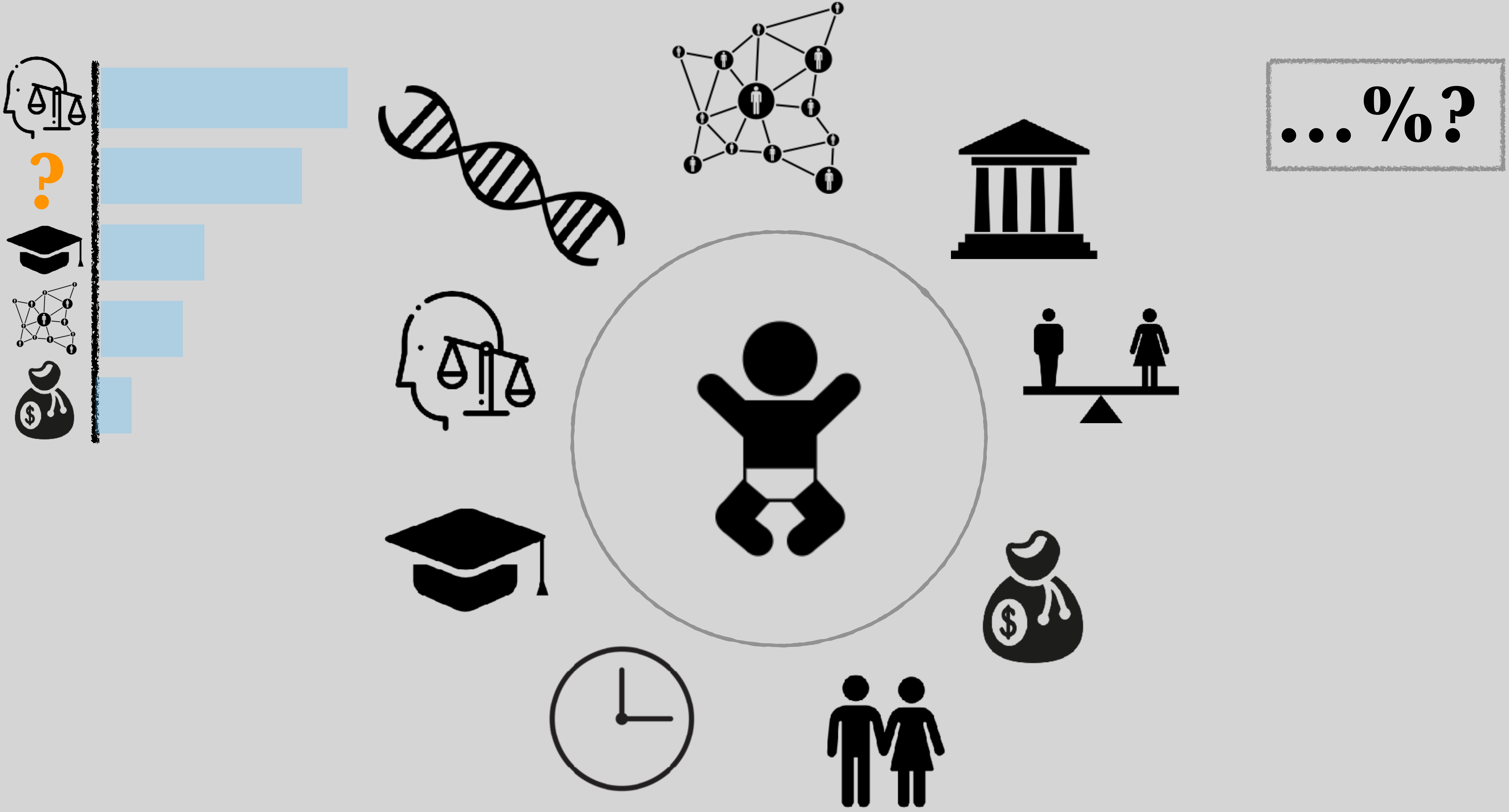
Progress in accuracy over time for ImageNet.

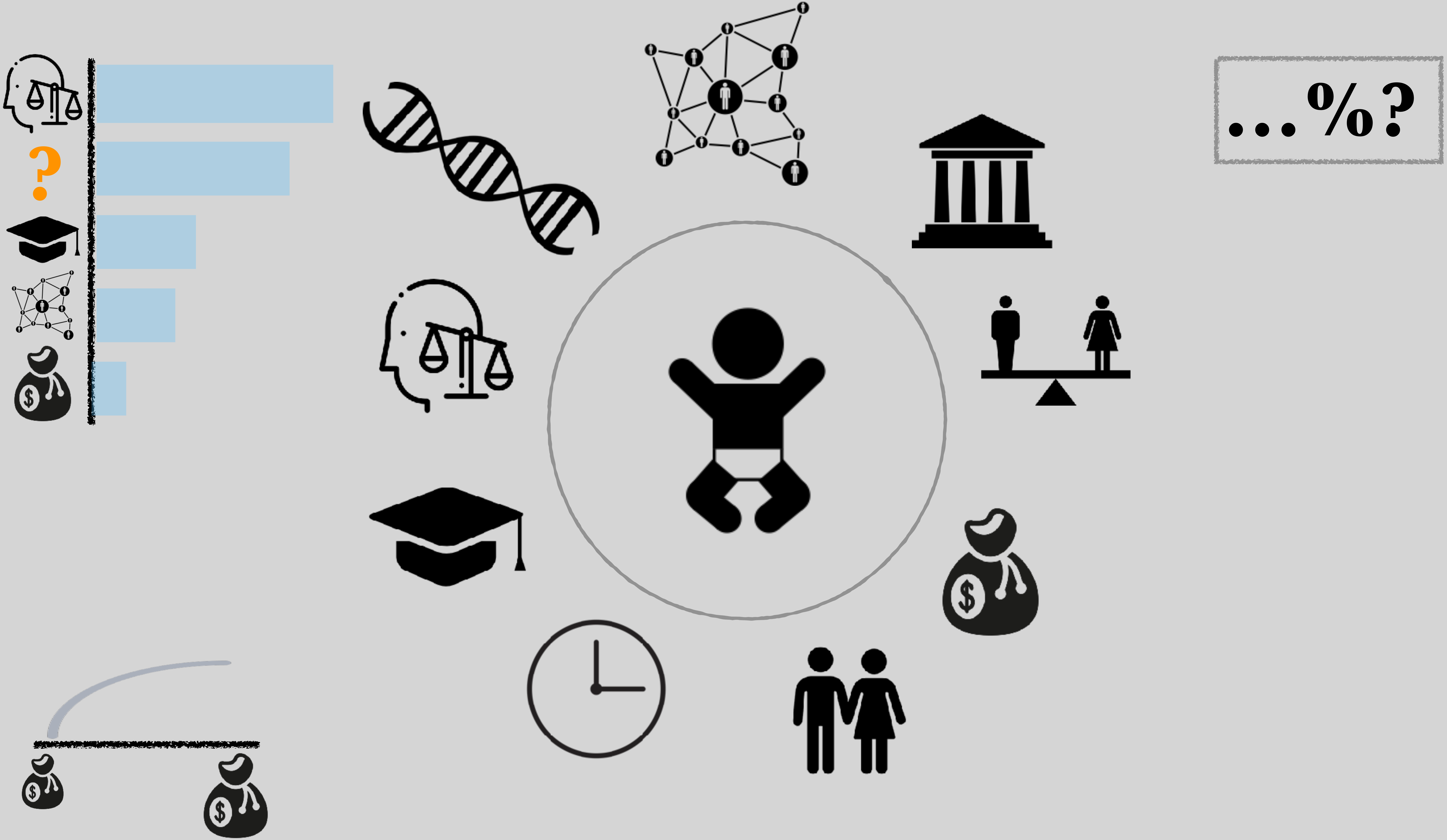
From: Research community dynamics behind popular AI benchmarks





…%?





PreFer

Predicting Fertility data challenge

PreFer

Predicting Fertility data challenge

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

Common Task

To predict who will have a child in the next three years

Outcome ['21-'23]

Background data
[data from 1995/2007 up to 2020]
[ages 18-45]

70% **TRAIN**

30% **HOLD-OUT**

Common Task

Rationale

- Difficult!
  [minimal policy-relevant test]
- Parity-specific
- Data availability
  [longer prediction timespan means fewer background data]

To predict who will have a child in the next three years

Outcome ['21-'23]

Background data
[data from 1995/2007 up to 2020]
[ages 18-45]

70% **TRAIN**

30% **HOLD-OUT**

Common Data

1 LISS panel

panel survey [2007 - current]

~1200 cases

'objective' and 'subjective' measures

15 waves x 10 core surveys
1000s variables

2 Social Statistics Netherlands

register data [1995 - current]

6 milion cases

'objective' measures

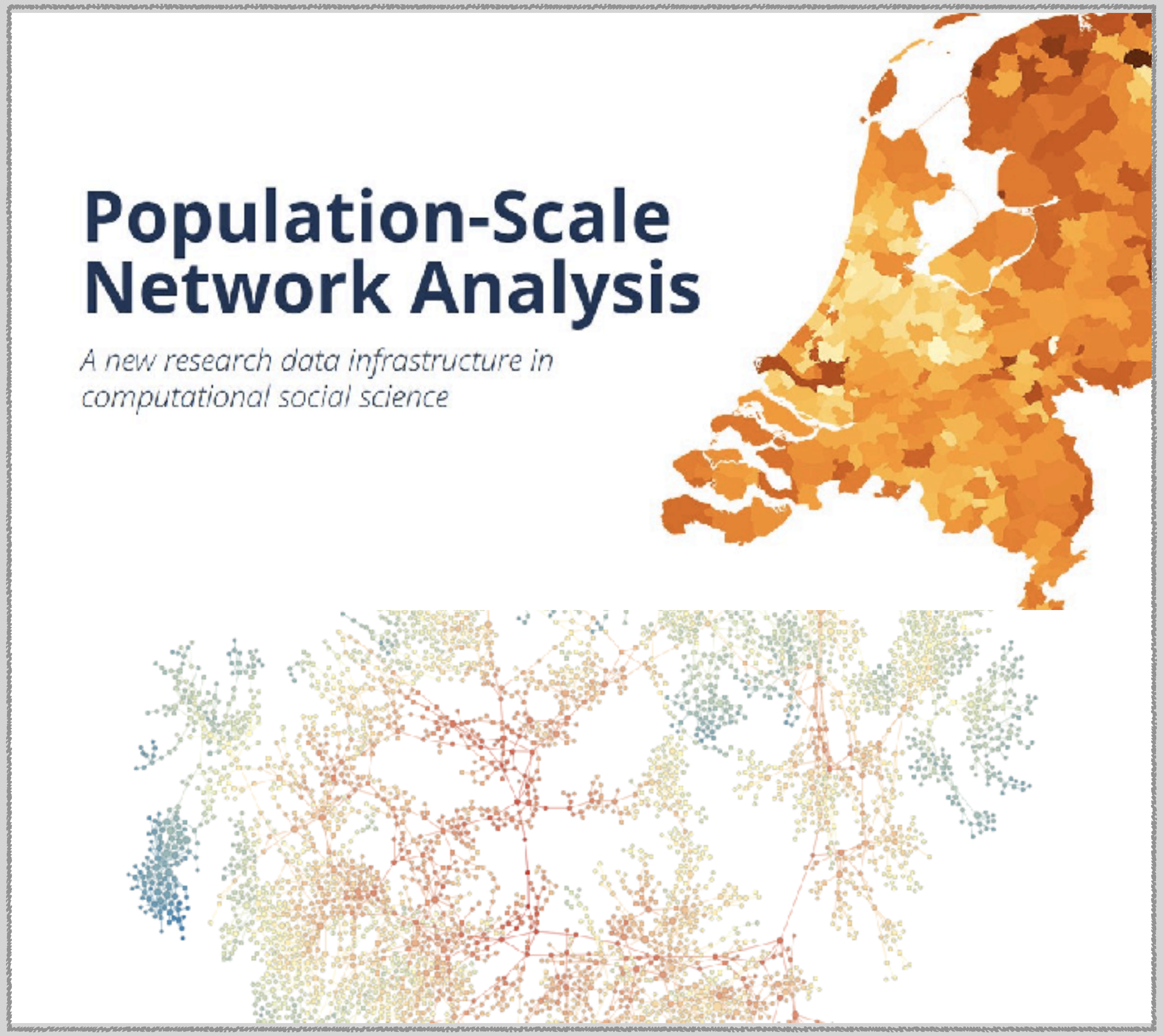
100s variables
(10000s variables?)

Common Data

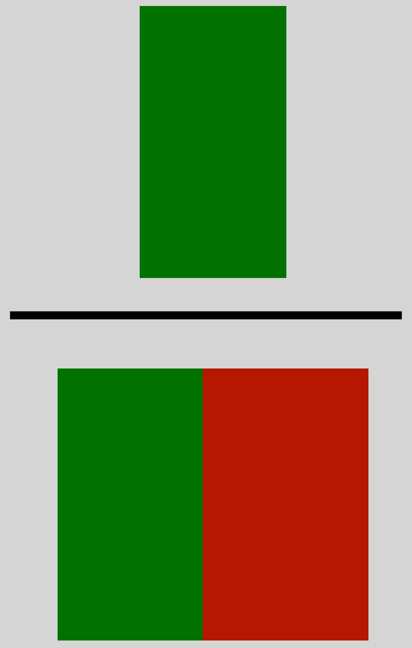
1 LISS panel

Background variables
Health
Religion and Ethnicity
Social Integration and Leisure
Family and Household
Work and Schooling
Personality
Politics and Values
Economic Situation:
  Assets, Income, Housing

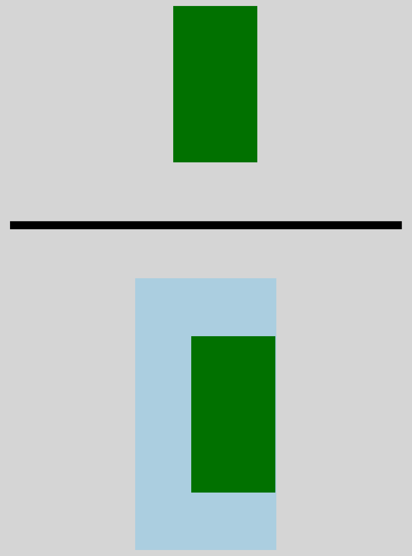
2 Social Statistics Netherlands



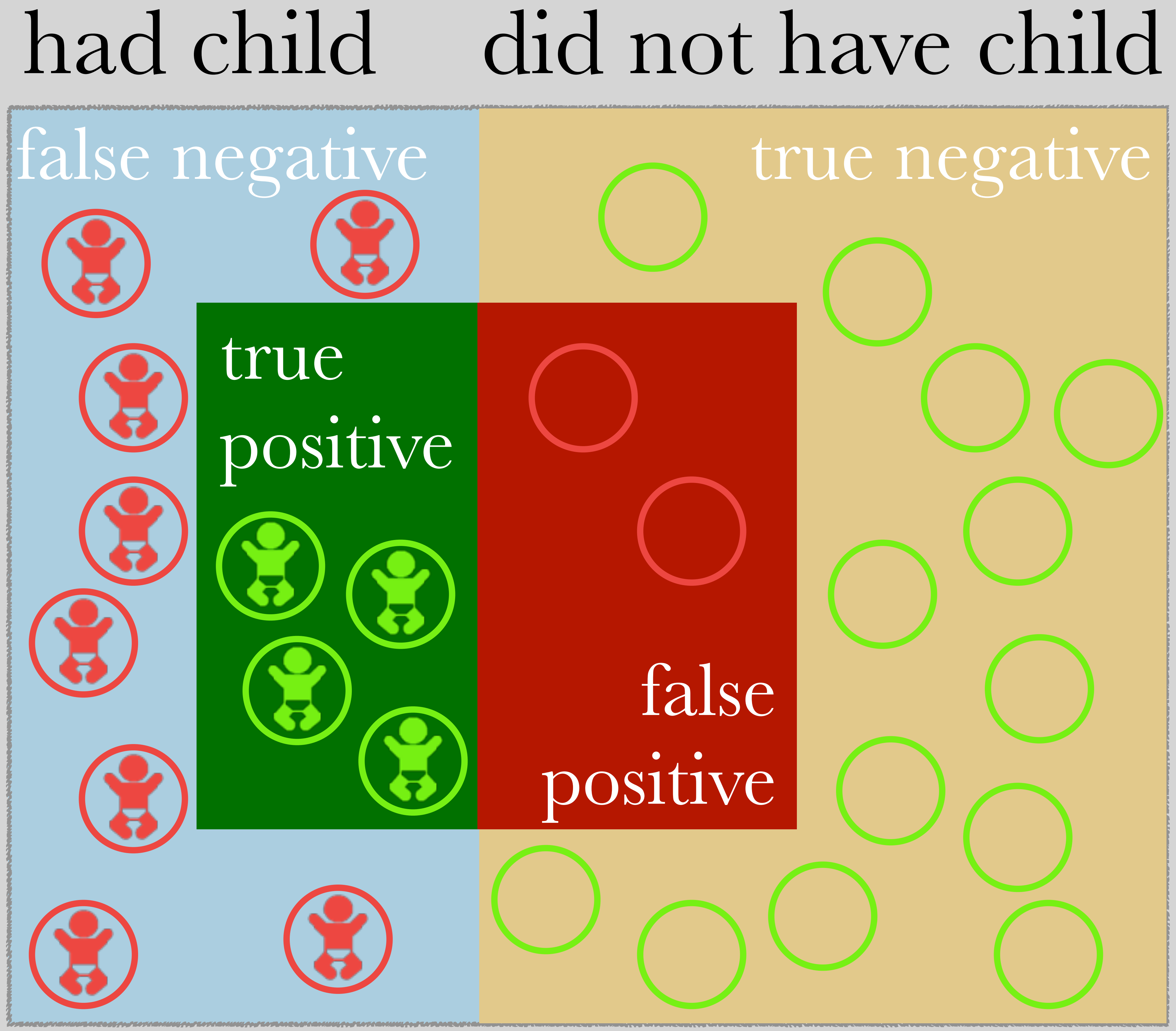
Common Metric: F1

precision = 

among those predicted to have child, % who had child

recall = 

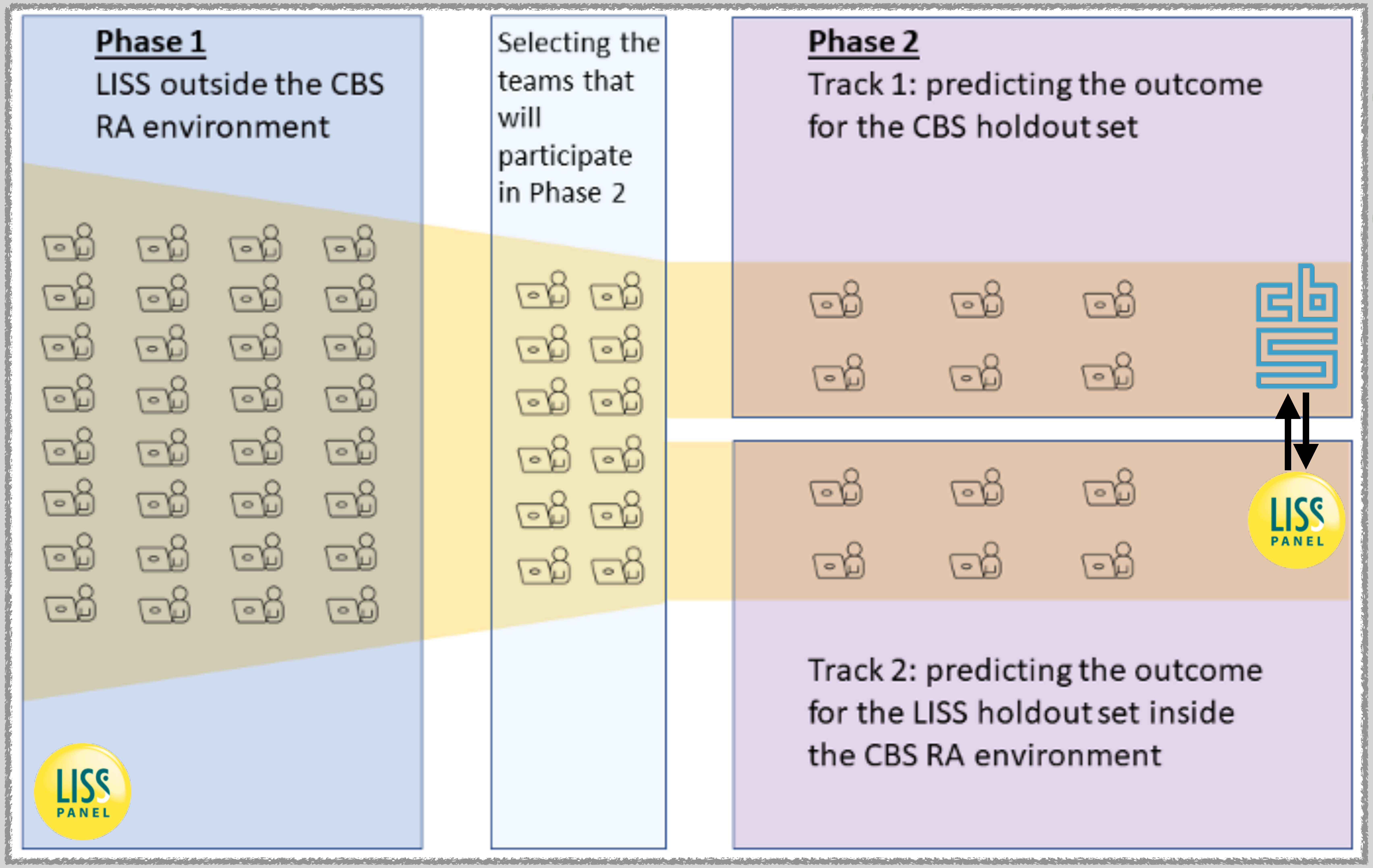
among those who had child, % predicted to have child



harmonic mean of precision and recall:

$$F1 = 2\,\frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

Overview

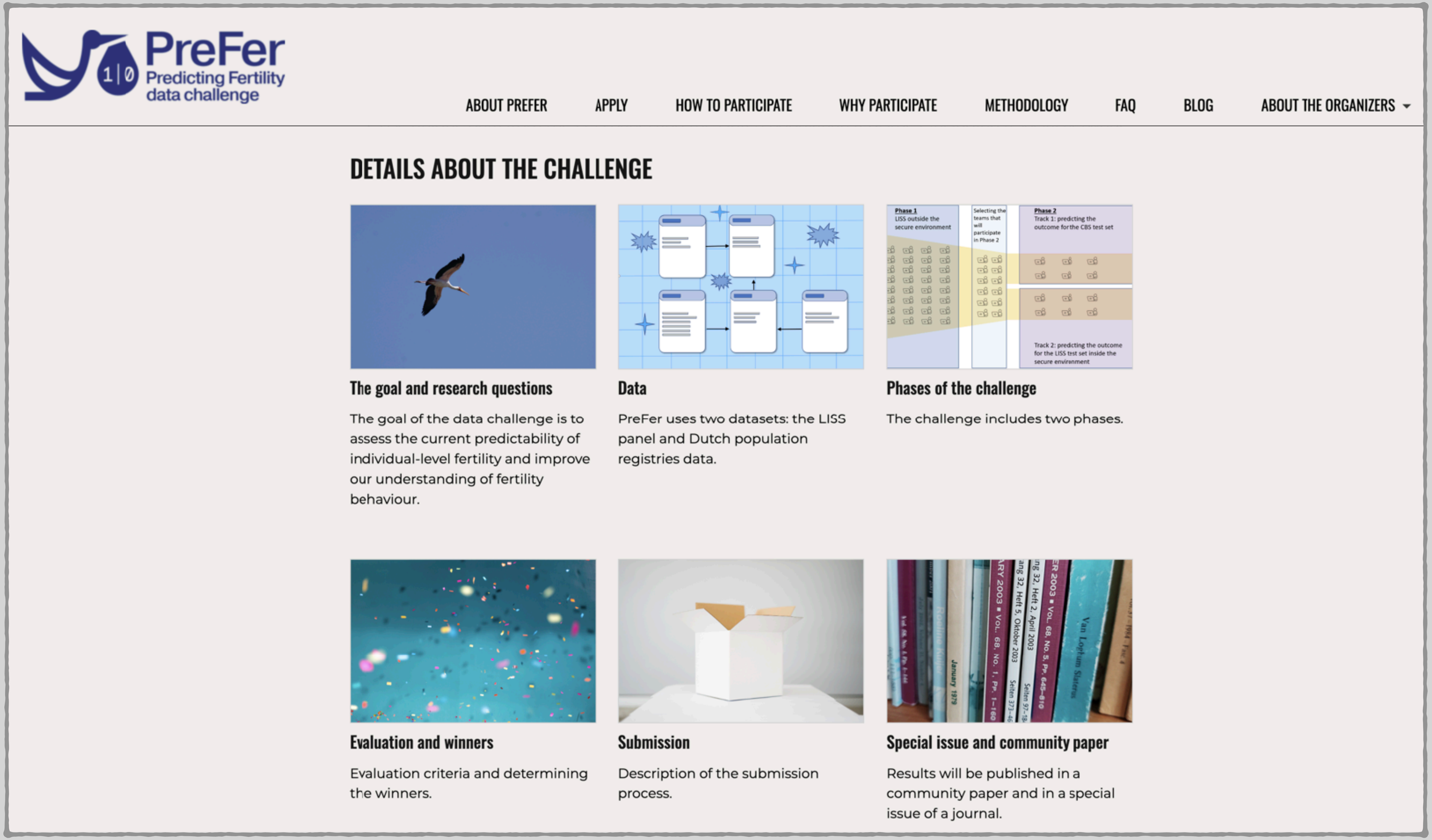


Evaluation criteria:

- F1 score *[3 winners]*
- Qualitative criteria *[2 winners]*
  - innovativeness: novel approach from social- or data science
  - improving understanding: what have we learned about fertility

Eyra

preferdatachallenge.nl



Timeline

January-March 2024	April-May 2024	June-September 2024	October 2024	2024 - 2025
Application	**Phase 1**	**Phase 2**	**Evaluation**	**Analysis**
Sign up to get notified when the application opens	Predict the outcome for the LISS holdout set	Predict the outcome for the LISS holdout set or/and CBS holdout set inside CBS RA	Evaluating the submissions and announcing the winners	Analyzing submitted methods, preparing publications

select winner phase 1 &
teams for phase 2

Who Can Participate?

1 LISS panel

No restrictions, but:
Access after data agreement

Data on your own computer

2 Social Statistics Netherlands

Restricted access
Only after vetting procedure

Remote secure environment

No uploads/downloads possible

Only available from within European Union + selected countries

Why Participate?

- eternal glory
- talk at and paid-for-trip to conference in exotic Netherlands

- publish paper [special issue]
- contribute to fertility research / computational social science
- work with amazing data
- test your ML skills and favourite algorithms

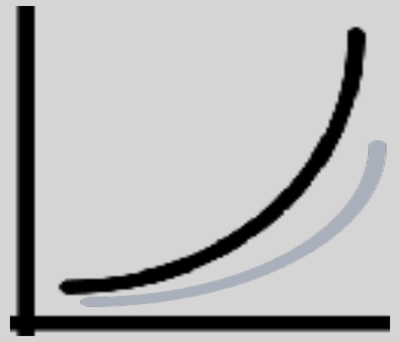
- students: learn new skills
- use in/for teaching

What Can We Learn?

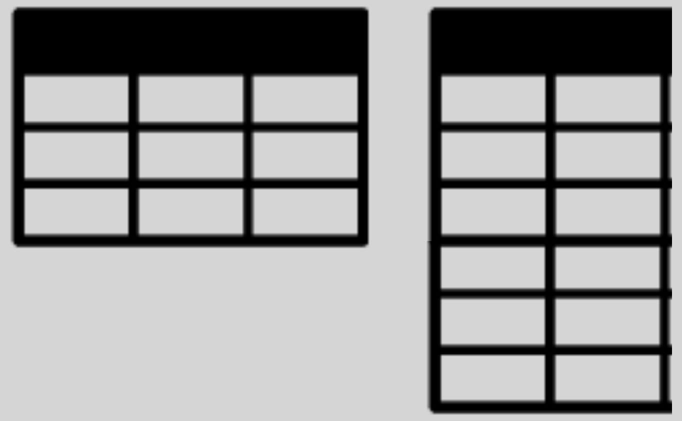
Science

Establish predictive ability and set benchmarks

Novel understanding through e.g. non-linearity, interactions



Scale versus scope, long versus wide



“subjective” versus “objective” measures

transfer learning **1+1=3**

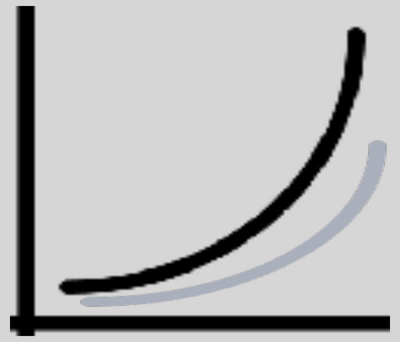
success further in the future **2025?**

What Can We Learn?

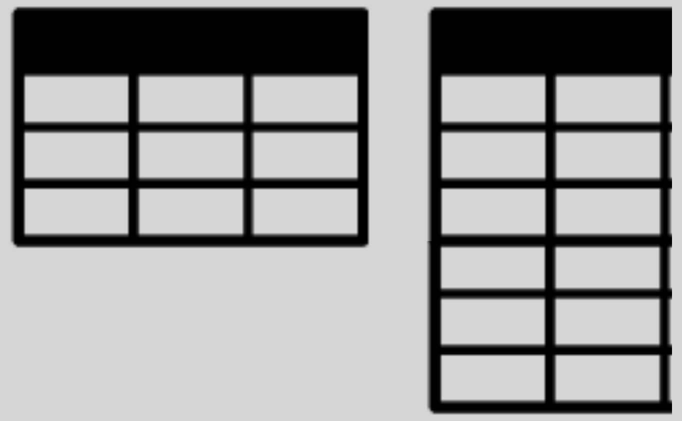
Science

Establish predictive ability and set benchmarks

Novel understanding through e.g. non-linearity, interactions



Scale versus scope, long versus wide



“subjective” versus “objective” measures

transfer learning **1+1=3**

success further in the future **2025?**

Policy

Debate on using intentions in forecasting

Quantifying unmet needs

PreFer
Predicting Fertility data challenge

- Be a part of a unique data challenge
- Contribute to fertility research & computational social sciences
- Publish research
- Work with amazing data:
  - LISS panel
  - Dutch population registries

SIGN UP HERE!

preferdatachallenge.nl

g.stulp@rug.nl