

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 Comptonⁱ, Debanjan Datta^j, Thomas Davidson^k, Anna Filippova^l, Connor Gilroy^m, Brian J. Goodeⁿ, 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. Stancescu^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 Chungⁱⁱ, William Eggert^{jj}, Gregory Faletto^{kk}, Zhilin Fan^{ll}, Jeremy Freese^{gg}, Tejomay Gadgil^{mm}, Josh Gagné^{gg}, Yue Gaoⁿⁿ, 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 Wangⁱⁱⁱ, Zhi Wang^{hhh,jjj}, Julia Wang^y, Flora Wang^{fff}, Samantha Weissman^y, Kirstie Whitaker^{kkk,lll}, Maria K. Wolters^{mmm}, Wei Lee Woonⁿⁿⁿ, 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^{vii}, 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}

Contributed by Sara McLanahan, January 24, 2020 (sent for review October 1, 2019; reviewed by Sendhil Mullainathan and Brian Uzzi)

How predictable are life trajectories? We investigated this question with a scientific mass collaboration using the common task method; 160 teams built predictive models for six life outcomes using data from the Fragile Families and Child Wellbeing Study, a high-quality birth cohort study. Despite using a rich dataset and applying machine-learning methods optimized for prediction, the best predictions were not very accurate and were only slightly better than those from a simple benchmark model. Within each outcome, prediction error was strongly associated with the family being predicted and weakly associated with the technique used to generate the prediction. Overall, these results suggest practical limits to the predictability of life outcomes in some settings and illustrate the value of mass collaborations in the social sciences.

life course | prediction | machine learning | mass collaboration

Social scientists studying the life course have described social patterns, theorized factors that shape outcomes, and estimated causal effects. Although this research has advanced scientific understanding and informed policy interventions, it is unclear how much it translates into an ability to predict individual life outcomes. Assessing predictability is important for three reasons. First, accurate predictions can be used to target assistance to children and families at risk (1, 2). Second, predictability of a life outcome from a person's life trajectory can indicate social rigidity (3), and efforts to understand differences in predictability across social contexts can stimulate scientific discovery and improve policy-making (4). Finally, efforts to improve predictive performance can spark developments in theory and methods (5).

In order to measure the predictability of life outcomes for children, parents, and households, we created a scientific mass collaboration. Our mass collaboration—the Fragile Families Challenge—used a research design common in machine learning but not yet common in the social sciences: the common task method (6). To create a project using the common task method, an organizer designs a prediction task and then recruits a large, diverse group of researchers who complete the task by predicting the exact same outcomes using the exact same data. These pre-

dictions are then evaluated with the exact same error metric that exclusively assesses their ability to predict held-out data: data that are held by the organizer and not available to participants. Although the structure of the prediction task is completely standardized, participants are free to use any technique to generate predictions.

The common task method produces credible estimates of predictability because of its design. If predictability is higher than expected, the results cannot be dismissed because of concerns about overfitting (7) or researcher degrees of freedom (8). Alternatively, if predictability is lower than expected, the results cannot be dismissed because of concerns about the limitations

Author contributions: M.J.S., I.L., A.T.K., J.B.-G., B.E.E., M.H., K.L., A.N., B.M.S., D.J.W., and S. McLanahan designed research; M.J.S., I.L., A.T.K., C.E.A., K.A.-G., A. Almaatouq, D.M.A., J.E.B., N.B.C., R.J.C., D.D., T.D., A.F., C.G., B.J.G., E.J., R.K., A. Kirchner, S. McKay, A.C.M., A.P., K.P., L.R., D.E.R., C.V.R., D.M.S., Y.S., A.U., E.H.W., M.A., A. Alhajri, B.A., R.A., R.B.A., L.P.A., L.B.-B., M.B., B.-R.C., W.E., G.F., Z.F., J.F., T.G., J.G., Y.G., A.H.-M., S.P.H., S.H., G.H., K.H., B.H., I.M.H., L.M.H., N.J., K.J., D.J., P.K., A. Karapetyan, E.H.K., B.L., N.L., M. Möser, A.E.M., M. Mahajan, N.M., H.M., D.M.-G., V.M., K.M.-G., A.M., Q.N., W.N., H.O., A.O., K.O., K.M.P., E.P., K.E.P., C.Q., T.R., A.S., T.S., L.S., B. Schonfeld, B. Sender, J.D.T., E.T., A.v.L., O.V., X.W., Z.W., J. Wang, F.W., S.W., K.W., M.K.W., W.L.W., J. Wu, C.W., K.Y., J.Y., B.Z., and C.Z. analyzed data; and M.J.S., I.L., A.T.K., D.K., and S. McLanahan wrote the paper.

Reviewers: S.M., University of Chicago; and B.U., Northwestern University.

Competing interest statement: B.E.E. is on the scientific advisory boards of Celsius Therapeutics and Freenome, is currently employed by Genomics plc and Freenome, and is on a year leave-of-absence from Princeton University.

This open access article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](#).

Data deposition: Predictions, code, and narrative explanations for valid submissions to the Fragile Families Challenge and code to reproduce the results of this paper are available from Dataverse at <https://doi.org/10.7910/DVN/CXSECU>. Data used in the Fragile Families Challenge are currently available to approved researchers from the Princeton University Office of Population Research Data Archive at <https://opr.princeton.edu/archive/>.

See [online](#) for related content such as Commentaries.

¹To whom correspondence may be addressed. Email: mjs3@princeton.edu or mclanaha@princeton.edu.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1915006117/-DCSupplemental>.

First published March 30, 2020.