The Unpredictability of Individual-Level Longevity

Flash — Socioeconomic Inequalities in Mortality

Casey F. Breen [1] Nathan Seltzer [1]

[1]University of California, Berkeley | Department of Demography

April 13, 2023

Social Science is increasingly interested in individual-level outcomes

- Researchers are increasingly seeking to pose and answer research questions about prediction at the individual-level (Hofman, Sharma and Watts, 2017; Salganik et al., 2020; Arpino, Le Moglie and Mencarini, 2022)

Social Science is increasingly interested in individual-level outcomes

- Researchers are increasingly seeking to pose and answer research questions about prediction at the individual-level (Hofman, Sharma and Watts, 2017; Salganik et al., 2020; Arpino, Le Moglie and Mencarini, 2022)
- Explosion in types and volume of data available + advances in computing have opened up new opportunities for prediction

Research question: how accurately can age of death be predicted from sociodemographic characteristics?

Research question: how accurately can age of death be predicted from sociodemographic characteristics?

Why should we care?

- How much do theories about mortality (e.g., fundamental cause theory) tell us about **individual-level longevity**?

Research question: how accurately can age of death be predicted from sociodemographic characteristics?

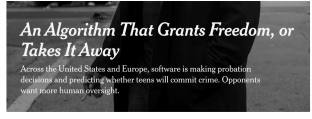
Why should we care?

- How much do theories about mortality (e.g., fundamental cause theory) tell us about **individual-level longevity**?
- Predictive algorithms are increasingly used for policy

Research question: how accurately can age of death be predicted from sociodemographic characteristics?

Why should we care?

- How much do theories about mortality (e.g., fundamental cause theory) tell us about **individual-level longevity**?
- Predictive algorithms are increasingly used for policy



How accurately can we predict individual-level longevity?

Our approach:

- Use large-scale, linked census + mortality data from CenSoc-DMF (Goldstein et al., 2021)

How accurately can we predict individual-level longevity?

Our approach:

- Use large-scale, linked census + mortality data from CenSoc-DMF (Goldstein et al., 2021)
  - Focus on cohort of 1910, age 29/30 when observed in 1940 Census (N = 130k)

How accurately can we predict individual-level longevity?

Our approach:

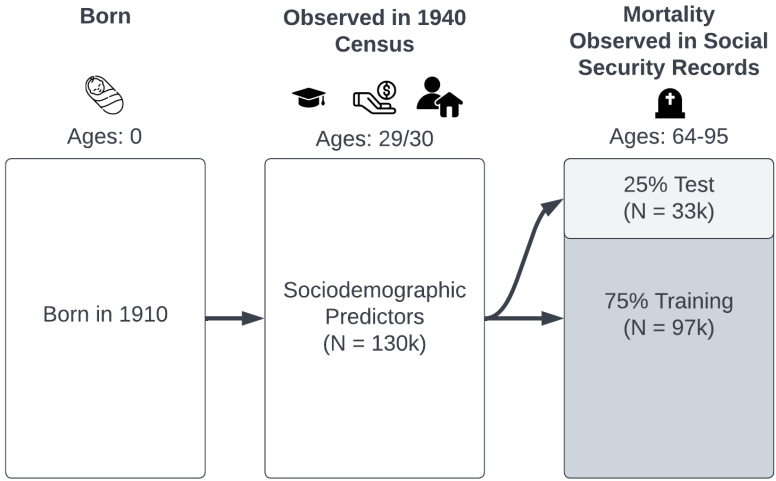
- Use large-scale, linked census + mortality data from CenSoc-DMF (Goldstein et al., 2021)
  - Focus on cohort of 1910, age 29/30 when observed in 1940 Census (N = 130k)
- Train 8 different machine learning algorithms to predict individual-level age of death

How accurately can we predict individual-level longevity?

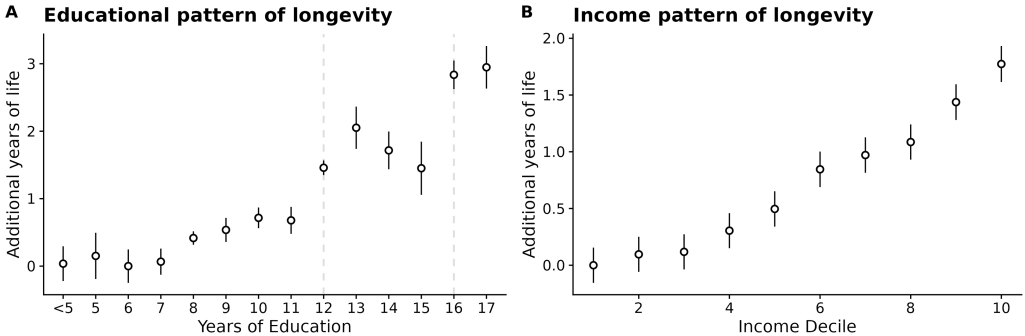
Our approach:

- Use large-scale, linked census + mortality data from CenSoc-DMF (Goldstein et al., 2021)
  - Focus on cohort of 1910, age 29/30 when observed in 1940 Census (N = 130k)
- Train 8 different machine learning algorithms to predict individual-level age of death

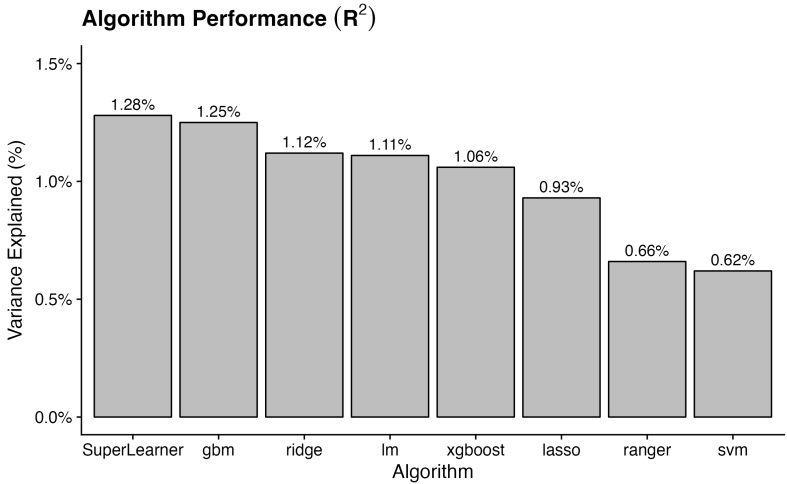
Train/test split for prediction exercise



CenSoc allows us to zoom in on "high-resolution" aggregate mortality disparities (e.g., education staircase)



Our best model only explains 1.3% of variation in age of death in test dataset



Conclusions

- We **cannot** use sociodemographic characteristics alone to accurately predict individual-level longevity

Conclusions

- We **cannot** use sociodemographic characteristics alone to accurately predict individual-level longevity
- Three takeaways:

Conclusions

- We **cannot** use sociodemographic characteristics alone to accurately predict individual-level longevity
- Three takeaways:
  1. We can still **study differences between groups**

Conclusions

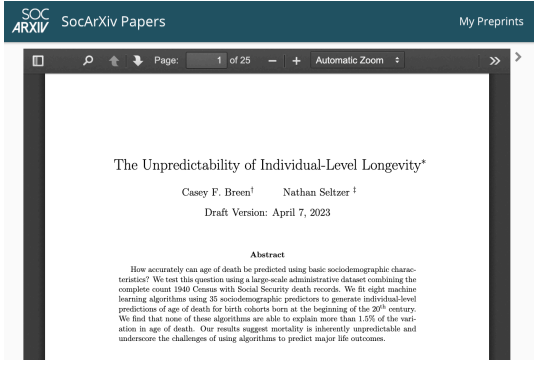
- We **cannot** use sociodemographic characteristics alone to accurately predict individual-level longevity
- Three takeaways:
  1. We can still **study differences between groups**
  2. Machine learning gives us modest gains over simple baseline model

Conclusions

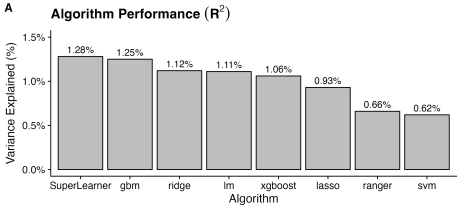
- We **cannot** use sociodemographic characteristics alone to accurately predict individual-level longevity
- Three takeaways:
  1. We can still **study differences between groups**
  2. Machine learning gives us modest gains over simple baseline model
  3. Healthy skepticism around using prediction for policy

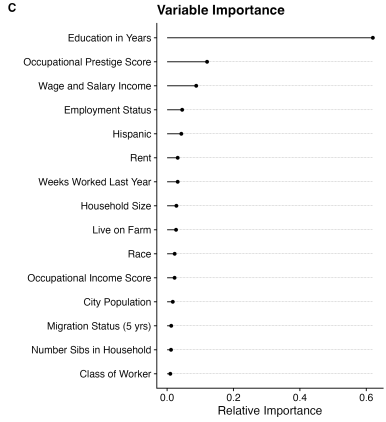
Working Paper on SocArXiv

- We would love feedback



Thank You





caseyfbreen

caseybreen@berkeley.edu

References

Arpino, Bruno, Marco Le Moglie and Letizia Mencarini. 2022. "What Tears Couples Apart: A Machine Learning Analysis of Union Dissolution in Germany." *Demography* 59(1):161–186.

Goldstein, J. R., M. Alexander, C. Breen, A. Miranda González, F. Menares, M. Osborne, M. Snyder and U. Yildirim. 2021. "Censoc Project." *CenSoc Mortality File: Version 2.0. Berkeley: University of California* .

Hofman, Jake M., Amit Sharma and Duncan J. Watts. 2017. "Prediction and Explanation in Social Systems." *Science* 355(6324):486–488.

Salganik, Matthew J., Ian Lundberg, Alexander T. Kindel, Caitlin E. Ahearn, Khaled Al-Ghoneim, Abdullah Almaatouq, Drew M. Altschul, Jennie E. Brand, Nicole Bohme Carnegie, Ryan James Compton, Debanjan Datta, Thomas Davidson, Anna Filippova, Connor Gilroy, Brian J. Goode, Eaman Jahani, Ridhi Kashyap, Antje Kirchner, Stephen McKay, Allison C. Morgan, Alex Pentland, Kivan Polimis, Louis Raes, Daniel E. Rigobon, Claudia V. Roberts, Diana M. Stanescu, Yoshihiko Suhara, Adaner Usmani, Erik H. Wang, Muna Adem, Abdulla Alhajri, Bedoor AlShebli, Redwane Amin, Ryan B. Amos, Lisa P. Argyle, Livia Baer-Bositis, Moritz Büchi, Bo-Ryehn Chung, William Eggert, Gregory Faletto, Zhilin Fan, Jeremy Freese, Tejomay Gadgil, Josh Gagné, Yue Gao, Andrew Halpern-Manners, Sonia P. Hashim, Sonia Hausen, Guanhua He, Kimberly Higuera, Bernie Hogan, Ilana M. Horwitz, Lisa M. Hummel, Naman Jain, Kun Jin, David Jurgens, Patrick Kaminski, Areg Karapetyan, E. H. Kim, Ben Leizman, Naijia Liu, Malte Möser, Andrew E. Mack, Mayank Mahajan, Noah Mandell, Helge Marahrens, Diana Mercado-Garcia, Viola Mocz, Katariina Mueller-Gastell, Ahmed Musse, Qiankun Niu, William Nowak, Hamidreza Omidvar, Andrew Or, Karen Ouyang, Katy M. Pinto, Ethan Porter, Kristin E. Porter, Crystal Qian, Tamkinat Rauf, Anahit Sargsyan, Thomas Schaffner, Landon Schnabel, Bryan Schonfeld, Ben Sender, Jonathan D. Tang, Emma Tsurkov, Austin van Loon, Onur Varol, Xiafei Wang, Zhi Wang, Julia Wang, Flora Wang, Samantha Weissman, Kirstie Whitaker, Maria K. Wolters, Wei Lee Woon, James Wu, Catherine Wu, Kengran Yang, Jingwen Yin, Bingyu Zhao, Chenyun Zhu, Jeanne Brooks-Gunn, Barbara E. Engelhardt, Moritz Hardt, Dean Knox, Karen Levy, Arvind Narayanan, Brandon M. Stewart, Duncan J. Watts and Sara McLanahan. 2020. "Measuring the Predictability of Life Outcomes with a Scientific Mass Collaboration." *Proceedings of the National Academy of Sciences* 117(15):8398–8403.