



Distinguishing Death from Disenrollment in Claims Data:

A readily implemented
machine learning algorithm

Jessica C. Young, Frank Yoon,
Nabarun Dasgupta, Debra E. Irwin,
Cory Pack, Toska Cooper, Shalu Shiv,
Sarah Bloemers, Teresa Gibson



THE CECIL G. SHEPS CENTER FOR
HEALTH SERVICES RESEARCH

Disclosures

- This work is supported by the US Food and Drug Administration (HHSF223201810183C).
- JY is supported by a National Research Service Award Post-Doctoral Traineeship from the Agency for Healthcare Research and Quality sponsored by The Cecil G. Sheps Center for Health Services Research, The University of North Carolina at Chapel Hill, Grant No. T32-HS000032.
- JY is a consultant for CERobs Consulting, LLC.



Distinguishing Death from Disenrollment

Background:

- Inability to identify dates of death in insurance claims is a major limitation
- If not the outcome of interest, death is a competing risk and poses a threat to validity when treated as non-informative right censoring
- Requiring survival leads to selection bias

Objective:

- Develop a user-friendly public algorithm to predict when disenrollment from an administrative claims database is indicative of death

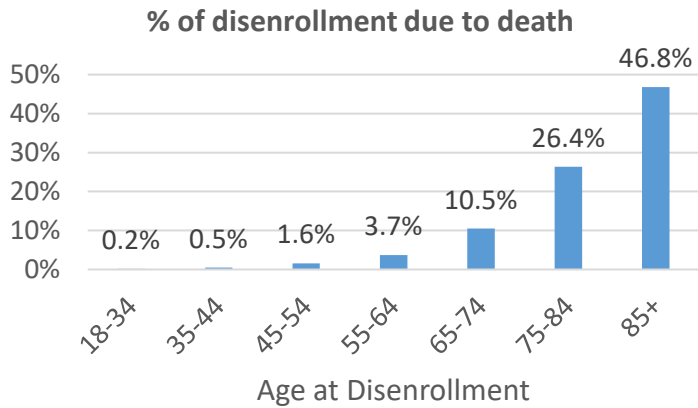
Data Sources:

- Administrative claims, 2007 – 2018
- 3 Sources of Death Data: inpatient discharge status, Social Security Death Index, data contributor provided death indicators

Machine Learning:

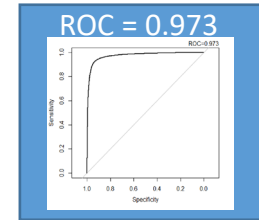
- Elastic-net regression w/ Synthetic Minority Oversampling TEchnique
- Partitioned data: 80% training, 20% testing
- Predicted outcome: Death date w/in 61 days of disenrollment

Distinguishing Death from Disenrollment



Candidate Predictors:

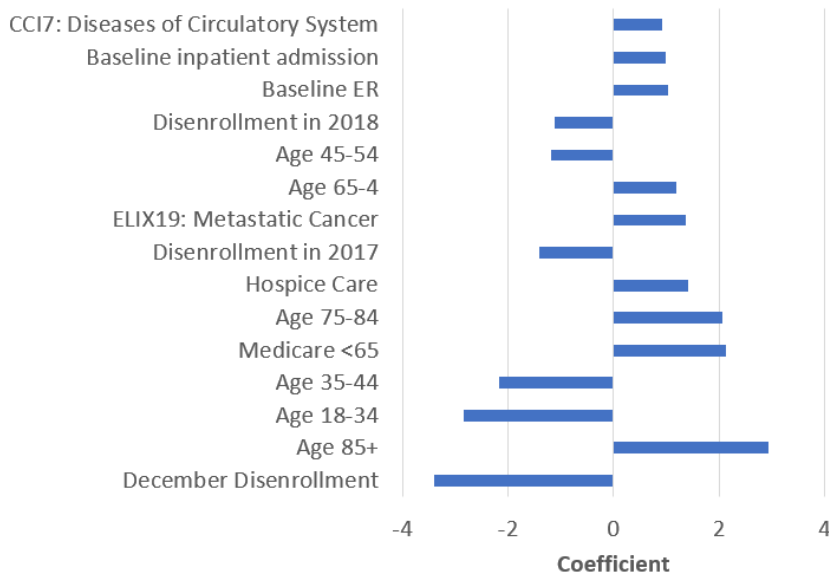
- Elixhauser comorbidity index
- HCUP Chronic Condition Indicator
- Age, sex, region
- Medicare eligibility
- Hospice, inpatient, and ER utilization
- Healthcare costs
- Month of disenrollment



Algorithm performance across predicted probability thresholds

Predicted Probability of Death	Sensitivity	Specificity
0.5	0.916	0.943
0.9	0.721	0.986

Estimated Coefficients - Selected Predictors



Example Patients



Female, 34 yrs, North Central
Complications of pregnancy
ER admission
July 2017 Disenrollment
Predicted Probability of Death: 0.01



Male, 77 yrs, West
PVD, complicated hypertension,
complicated diabetes
Elix score = 2
ER and IP admission
May 2010 Disenrollment
Predicted Probability of Death: 0.98

Thank You!

Questions / Comments:
Jessica.Young@unc.edu

Read more here:

[Machine Learning Can Unlock Insights
Into Mortality](#)

American Journal of Public
Health **111**, S65_S68, <https://doi.org/10.2105/AJPH.2021.306418>



SCAN ME



THE CECIL G. SHEPS CENTER FOR
HEALTH SERVICES RESEARCH