BaBE: Enhancing Fairness via Estimation of Latent Explaining Variables

Introduction

Age is an important factor in determining retirement time, screening time for oncologic and cardiovascular diseases, and similar decisions. However, a chronological age (the time from birth to the current moment) does not always coincide with a biological age (an estimate based on biological markers that indicate the cellular age of an organism). Moreover, the error between biological age and chronological estimate is disparate for black and white populations. Recent studies show that black people on average are older biologically than chronologically, while white people have lower biological age compared to their chronological age [4, 1, 2].

The Data Set

The National Health and Nutrition Examination Survey (NHANES) is a series of studies that are intended to evaluate the health and nutritional status of adults and children in the United States. The survey is unique in that it incorporates in-depth interviews and detailed physical examinations. Health-related questions and demographics are included in the NHANES interview. For the survey, the sample was selected to represent the US population of all ages. To produce reliable statistics, NHANES oversamples individuals aged 60 and over, African Americans, and Hispanics. The National Health and Nutrition Examination Survey (NHANES) [1] conducted by the National Center for Health Statistics is a popular source for studying biological aging [3, 7, 5, 6]. The data set consists of 8243 samples. For our experiments, we use three variables from the data set, race (black or white), which is out S, chronological age (20-90), which is our Z, and an estimate of the biological age of the original KDM [2] biological age (variable 'kdm0') which is our E. We choose chronological or biological age 75 or more as the threshold to set $Y_Z = 1$ and $Y_E = 1$. This age group shows the most racial disparity in biological aging in the NHANES data set (Figure 1). Additionally, it is a reasonable age to check for age-related diseases or consider retirement.

[1] $https://www.cdc.gov/nchs/nhanes/about_nhanes.htm$

[2] Klemera and Doubal's method for calculating the biological age from the set of biomarkers.

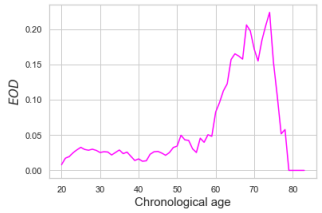

Figure 1: The graph shows the equal opportunity difference (EOD) between the Y_E and Y_Z, when different thresholds for Z (chronological age) are selected. The disparity is largest around the chronological age equal 75 years.

The Results of the Experiments

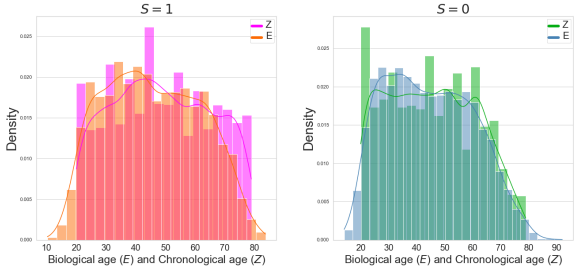

Figure 2: Distributions of E and Z for $S = 1$ (left) and $S = 0$ (right) in NHANES data set.

In this section, we show statistics and plots for the NHANES dataset. We applied BaBE using Method 2, because the conditional distribution of $Z|E, S$ does not allow the accurate estimation of every individual E. However, it still allows us to recover the aggregated distribution and estimate $\hat{Y}_{\hat{E}}$. Consistent with method 2 we report only EOD and $Acc(\hat{Y}_{\hat{E}}, Y_E)$, $Acc(Y_Z, Y_E)$.

Figure 3 shows the accuracy resulting from the application of BaBE, DI, and NB to the NHANES data set when the threshold is set to 75 years of age or older. BaBE achieves better overall accuracy and significantly better accuracy for $S = 1$.

Figure 4 shows the equal opportunity from the application of BaBE, DI, and NB to the NHANES data set. BaBE achieves EOD close to zero. DI and NB preprocessing methods do not differ significantly from the estimated EOD considering the original Z. DI and NB are designed to aim for a statistical disparity that is equal to zero. The statistical disparity in the NHANES data is very small (owing to the oversampling of the minority population), so much preprocessing is not needed if the goal is optimizing statistical parity.

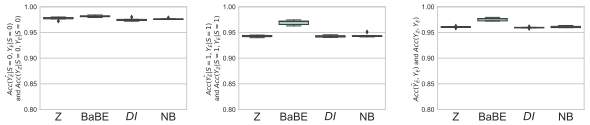

Figure 3: Experiments on the NHANES data. The accuracy between $\hat{Y}_Z$ and Y_E (for Z), and between $\hat{Y}_{\hat{E}}$ and Y_E.

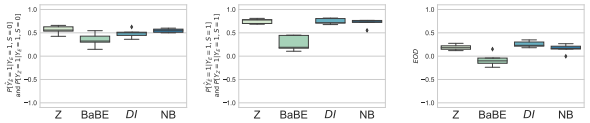

Figure 4: Experiments on the NHANES data. EOD.

References

[1] Farina, M. P.; Kim, J. K.; and Crimmins, E. M. 2023. Racial/ethnic differences in biological aging and their life course socioeconomic determinants: The 2016 Health and Retirement Study. *Journal of aging and health*, 35(3-4): 209–220.

[2] Graf, G. H.; Crowe, C. L.; Kothari, M.; Kwon, D.; Manly, J. J.; Turney, I. C.; Valeri, L.; and Belsky, D. W. 2022. Testing Black-White disparities in biological aging among older adults in the United States: analysis of DNA-methylation and blood-chemistry methods. *American journal of epidemiology*, 191(4): 613–625.

[3] Kwon, D.; and Belsky, D. W. 2021. A toolkit for quantification of biological age from blood chemistry and organ function test data: BioAge. *GeroScience*, 43: 2795–2808.

[4] Levine, M. E.; and Crimmins, E. M. 2014. Evidence of accelerated aging among African Americans and its implications for mortality. *Social Science & Medicine*, 118: 27–32.

[5] Liu, W.; Wang, J.; Wang, M.; Hou, H.; Ding, X.; Ma, L.; and Liu, M. 2023. Oxidative Stress Factors Mediate the Association Between Life's Essential 8 and Accelerated Phenotypic Aging: NHANES 2005-2018. *The Journals of Gerontology: Series A*, glad240.

[6] Nguyen, L.; Chon, J.; Kim, E.; Cheng, J.; and Ebersole, J. 2022. Biological aging and periodontal disease: analysis of NHANES (2001–2002). *JDR Clinical & Translational Research*, 7(2): 145–153.

[7] Xu, Y.; Wang, X.; Belsky, D. W.; McCall, W. V.; Liu, Y.; and Su, S. 2023. Blunted rest–activity circadian rhythm is associated with increased rate of biological aging: an analysis of NHANES 2011–2014. *The Journals of Gerontology: Series A*, 78(3): 407–413.