Project 3 Final report

Shuai Zhu

2024-11-27

Introduction

The National Health and Nutrition Examination Survey (NHANES) is a comprehensive program designed to evaluate the health and nutritional status of adults and children across the United States. As part of its mission, NHANES collects extensive data on various health determinants, including physical activity, chronic diseases, and mortality. This project utilizes the NHANES 2011-2014 dataset, focusing on physical activity metrics, mortality outcomes, and demographic factors. Specifically, the dataset includes accelerometer-derived measures of physical activity such as total MIMS (monitor-independent movement summary), L5 MIMS, and RA MIMS .

The primary objective of this study is to assess and rank the predictive performance of these physical activity metrics on mortality outcomes. The analysis considers both linear and non-linear associations between activity metrics and mortality and incorporates sex-stratified models to account for potential gender-specific differences in these relationships.

Method

Data Cleaning

Participants aged 40 to 80 years with at least three valid days of accelerometer data were included in the analysis. Further filtering to ensure complete cases across all relevant variables narrowed the dataset from an initial 19,141 participants to 5,564 participants. This cleaning process ensured that the analysis would not be confounded by missing data or insufficient physical activity records.

Data Analysis

Time-to-event methods were employed to evaluate the relationship between physical activity and mortality, accounting for potential non-linear associations. Cox proportional hazards models, both univariable and multivariable, were used to quantify these relationships. In the multivariable models, each physical activity metric was included alongside covariates such as BMI, age, and self-reported diagnoses of congestive heart failure and coronary heart disease. The non-linear associations were assessed using penalized splines (psplines) to capture potential complex patterns in the data.

Separate models were fitted for males and females to address possible gender-specific differences in the association between physical activity and mortality. Predictive performance of the models was evaluated using 10-fold cross-validation, with the concordance index (C-index) serving as the metric to rank their accuracy. All analyses were conducted using R version 4.3.3. The linear and non-linear Cox models were implemented using the `mgcv` package, and the concordance indices were calculated using the `survival` package.

Results

Table 1 summarizes the demographic characteristics and physical activity metrics for included and excluded participants, stratified by gender. Participants excluded from the analysis had fewer than three valid accelerometer days or missing values for physical activity and demographic variables. The distribution of total MIMS, L5 MIMS, and RA MIMS was similar across included and excluded groups, though the mortality rate was slightly higher among excluded participants. These small differences are unlikely to affect the validity of the Cox proportional hazards model.

Table 2 presents the ranking of univariable and multivariable Cox models, both linear and non-linear, for each physical activity metric, stratified by gender. Among males, multivariable models generally outperformed univariable models in terms of predictive accuracy. Across all models, total MIMS consistently demonstrated the highest predictive performance, followed by RA MIMS, with L5 MIMS performing the worst. The concordance indices of linear and non-linear models for each physical activity metric were comparable, suggesting that non-linear associations do not significantly improve predictive performance.

The results for females mirrored those observed in males. Total MIMS remained the most predictive metric, followed by RA MIMS and L5 MIMS. There was no evidence to suggest that incorporating non-linear splines enhanced the predictive performance of the models. These findings underscore the utility of total MIMS as the most robust physical activity measure for predicting mortality, with no added benefit from non-linear modeling approaches.

Conclusion

This study leveraged accelerometer-derived physical activity metrics from the NHANES 2011-2014 dataset to evaluate their predictive performance for mortality using both linear and non-linear Cox proportional hazards models. The findings highlight that total MIMS

consistently demonstrated the strongest predictive performance for mortality among both males and females, outperforming other activity metrics such as RA MIMS and L5 MIMS. Multivariable models generally provided better predictive accuracy than univariable models, suggesting the importance of adjusting for covariates such as BMI, age, and pre-existing conditions. Furthermore, the results indicated that non-linear modeling approaches using penalized splines did not significantly improve predictive performance compared to linear models.

Despite these findings, the study has several limitations. the study did not account for potential residual confounding from unmeasured factors such as dietary habits, socioeconomic status, or genetic predispositions, which could influence both physical activity and mortality. Lastly, while the stratification by gender addressed potential sex-specific differences, other intersectional factors, such as race or ethnicity, were not explored. Future research could address these limitations by incorporating additional demographic and behavioral variable. This would provide a more comprehensive understanding of the relationship between physical activity and mortality and improve the applicability of findings across diverse populations.

Table 1: Summary of NHANES dataset

	Include		Exclude	
	Male	Female	Male	Female
	(N=2636)	(N=2938)	(N=792)	(N=783)
Total MIMS				
Mean (SD)	12300 (3950)	13300 (3890)	10600 (4560)	12000 (5220)
Median [Min, Max]	12000 [295, 36600]	13200 [1070, 31500]	10600 [264, 32900]	12200 [321, 29500]
Missing	0 (0%)	0 (0%)	593 (74.9%)	579 (73.9%)
L5 MIMS				
Mean (SD)	1.10 (0.838)	0.955 (0.686)	1.06 (0.996)	0.915 (0.815)
Median [Min, Max]	0.839 [0, 8.28]	0.747 [0.000773, 5.67]	0.740 [0, 7.03]	0.667 [0, 4.94]
Missing	0 (0%)	0 (0%)	593 (74.9%)	579 (73.9%)
RA MIMS				
Mean (SD)	0.851 (0.0949)	0.878 (0.0780)	0.835 (0.117)	0.863 (0.114)
Median [Min, Max]	0.876 [0.275, 1.00]	0.901 [0.416, 0.999]	0.866 [0.253, 1.00]	0.894 [0.466, 1.00]
Missing	0 (0%)	0 (0%)	593 (74.9%)	579 (73.9%)
Age				
Mean (SD)	59.9 (12.1)	59.4 (12.1)	57.8 (11.9)	58.1 (12.8)
Median [Min, Max]	60.0 [40.0, 80.0]	59.0 [40.0, 80.0]	57.0 [40.0, 80.0]	56.0 [40.0, 80.0]
Number of good days				
Mean (SD)	6.47 (0.999)	6.43 (1.01)	3.11 (2.36)	3.06 (2.40)
Median [Min, Max]	7.00 [3.00, 7.00]	7.00 [3.00, 7.00]	2.00 [1.00, 7.00]	2.00 [1.00, 7.00]
Missing	0 (0%)	0 (0%)	593 (74.9%)	579 (73.9%)
Months of follow-up				
Mean (SD)	78.2 (20.5)	79.3 (19.0)	76.7 (21.7)	76.5 (21.2)
Median [Min, Max]	80.0 [0, 111]	80.0 [1.00, 110]	79.0 [2.00, 109]	79.0 [0, 109]
Missing	0 (0%)	0 (0%)	7 (0.9%)	7 (0.9%)
Diabetes				
No	2118 (80.3%)	2422 (82.4%)	652 (82.3%)	649 (82.9%)
Yes	518 (19.7%)	516 (17.6%)	138 (17.4%)	131 (16.7%)
Missing	0 (0%)	0 (0%)	2 (0.3%)	3 (0.4%)
congestive heart failure				
No	2503 (95.0%)	2801 (95.3%)	752 (94.9%)	732 (93.5%)
Yes	133 (5.0%)	137 (4.7%)	28 (3.5%)	42 (5.4%)
Missing	0 (0%)	0 (0%)	12 (1.5%)	9 (1.1%)
coronary heart disease				
No	2429 (92.1%)	2812 (95.7%)	725 (91.5%)	749 (95.7%)
Yes	207 (7.9%)	126 (4.3%)	47 (5.9%)	21 (2.7%)
Missing	0 (0%)	0 (0%)	20 (2.5%)	13 (1.7%)
Body mass index				
Normal	648 (24.6%)	768 (26.1%)	219 (27.7%)	214 (27.3%)
Underweight	32 (1.2%)	49 (1.7%)	11 (1.4%)	14 (1.8%)
Overweight	1056 (40.1%)	828 (28.2%)	298 (37.6%)	213 (27.2%)
Obese	900 (34.1%)	1293 (44.0%)	209 (26.4%)	272 (34.7%)
Missing	0 (0%)	0 (0%)	55 (6.9%)	70 (8.9%)
Mortality rate				
Mean (SD)	0.148 (0.355)	0.117 (0.321)	0.161 (0.367)	0.137 (0.344)
Median [Min, Max]	0 [0, 1.00]	0 [0, 1.00]	0 [0, 1.00]	0 [0, 1.00]
Missing	0 (0%)	0 (0%)	7 (0.9%)	7 (0.9%)

Table 2: Ranking of Concordance by male and female

	Male		Female	
Rank	Model	Concordance	Model	Concordance
1	Multi Non linear TMIMS	0.811	Multi TMIMS	0.808
2	Multi TMIMS	0.810	Multi Non linear TMIMS	0.808
3	Multi RA TMIMS	0.809	Multi RA TMIMS	0.804
4	Multi Non linear RA TMIMS	0.809	Multi Non linear RA TMIMS	0.803
5	Multi Non linear L5 MIMS	0.797	Multi L5 MIMS	0.791
6	Multi L5 MIMS	0.796	Multi Non linear L5 MIMS	0.791
7	Uni TMIMS	0.748	Uni TMIMS	0.743
8	Uni Non linear TMIMS	0.748	Uni Non linear TMIMS	0.743
9	Uni RA TMIMS	0.649	Uni RA TMIMS	0.636
10	Uni Non linear RA TMIMS	0.649	Uni Non linear RA TMIMS	0.636
11	Uni L5 MIMS	0.492	Uni L5 MIMS	0.483
12	Uni Non linear L5 MIMS	0.488	Uni Non linear L5 MIMS	0.483