Supplementary Material for: Physical activity and mobility differentially predict non-demented executive function change: Do sex and Alzheimer’s genetic risk moderate these associations?

## Participants

Participants were non-demented community dwelling older adults from the Victoria Longitudinal Study (VLS). The VLS is a large-scale, multi-faceted longitudinal investigation of human aging which incorporates examination of cognitive and health outcomes as influenced by multiple factors (Dixon & de Frias, 2004). The VLS follows a Longitudinal Sequential design. The first sample began in the late 1980s, the second sample began in the 1990s, and the third sample began in the early 2000s. Each sample is independent and tested repeatedly and regularly (approximately 4-year intervals). Sample 1 has been tested on nine different occasions (over ~25 years), Sample 2 has been tested on six different occasions (over ~20 years), and Sample 3 has been tested on four occasions (over ~12 years). As a focus of this study is to examine change in EF as predicted by genetic variant, participants were limited to a source subsample that provided biofluid for genotyping between 2009 and 2012 (*n* = 695), bridging all three main VLS core samples. Specifically, the present study consisted of Sample 1 (waves 6 and 7), Sample 2 (waves 4 and 5), and Sample 3 (waves 1, 2, and 3), for a total individualized duration of up to 9 years. Retention rates for each available and defined interval are as follows (a) S1 W1-W2 = 84%; (b) S2 W1-W2 = 83%; (c) S3 W1-W2 = 84%; (d) S3 W2-W3 = 86%; (e) S3 W1-W3 = 76%.

All present data collection procedures within the VLS are in full and certified compliance with prevailing human research ethics guidelines and boards. All participants provided informed written consent.

Exclusionary criteria were applied to the genotyped sub-sample to obtain a non-demented (to perform the executive tasks), non-mobility impaired (no movement disorders) sample with the assumed differential ability to participate in everyday physical activities and perform the gait and balance tasks in the lab. We evaluated exclusionary criteria based on associations with the main constructs of interest to our study. Specifically, severe conditions such a spinal condition, epilepsy, depression, or thyroid condition can impair the ability to participate in EPA, so were included as exclusionary criteria. Individuals with head injuries, drug or alcohol dependence, anti-psychotic medication use, or neurological conditions were also excluded to obtain a non-demented sample. Severe diabetes has been associated with lower EF performance (Nandipati, Luo, Schimming, Grossman, & Sano, 2012), greater cognitive decline in non-demented older adults (Yaffe et al., 2012), cerebrovascular disease pathology (Abner et al., 2016), and structural and functional brain changes outside of dementia (Moheet, Mangia, & Seaquist, 2015). Additionally, individuals with severe diabetes may experience gait disturbances or alterations (Brach, Talkowski, Strotmeyer, & Newman, 2008). Therefore, we chose to exclude participants with ratings of “severe” diabetes. High blood pressure is associated with gait slowing in well-functioning adults (Rosano et al., 2011), and poorer cognitive decline and impairment (Aronow, 2017). Low blood pressure has also been found to be associated with cognitive impairment (Novak & Hajjar, 2010). Therefore, we chose to exclude the participants with high and low blood pressure.

The final exclusionary criteria chosen were applied at baseline to obtain a non-demented subsample, physically able to participate in EPA, and without impairments in mobility: (a) MMSE score < 24 (n = 8), (b) “severe” ratings for a health condition (i.e., high blood pressure (n = 8), low blood pressure (n = 3), diabetes (n = 29), epilepsy (n = 1), thyroid conditions (n = 6), depression (n = 12), head injury (n = 12), spinal condition (n = 17)), (c) reported drug or alcohol dependence (n = 8), (d) reported use of anti-psychotic medications (n = 4), (e) self-reported “moderate” cases of neurological conditions, such as Parkinson’s or stroke (n = 4), and (f) insufficient EF, or mobility data (n = 44). A total of 156 participants were excluded.

## Statistical Analyses

Structural equation modeling (SEM) was conducted using Mplus 7 (Muthén & Muthén, 1998 - 2012). SEM compares a structural model, comprised of a system of regression equations, to empirical data. This approach allows analysis of latent variables and their relationships, concurrently and longitudinally, while incorporating the measurement error within the measurement model.

**Executive Function (EF).** Four neuropsychological measures were used to represent two dimensions of EF, with two measures each of shifting, and inhibition, (Miyake et al., 2000). All four measures have all been used in standard form with older adults in VLS studies reporting psychometric (Bielak, Mansueti, Strauss, & Dixon, 2006) structural and neuropsychological (de Frias, Dixon, & Strauss, 2006; de Frias, Dixon, & Strauss, 2009), genetic (Sapkota, Wiebe, Small, & Dixon, 2015), health (McFall et al., 2014; McFall et al., 2013), and lifestyle (de Frias & Dixon, 2014) factors.

*Shifting.* (1) Brixton Spatial Anticipation Test (Burgess & Shallice, 1997) indexed shifting by measuring rule-attainment. The Brixton test consists of a 56-page stimulus booklet, each page showing the same display of 10 circles, with each circle numbered from 1 to 10. On each page, one of the circles is filled with a blue color. The position of this filled circle changes from one page to the next. The changes in position are governed by a series of simple rules that vary without warning. Participants were required to predict the blue circle placement based on previous presented patterns. Responses were considered correct if the response followed the current pattern, or when the trials changed, if it would have been correct had the pattern remained the same. The total errors were recorded (on a maximum of 54 trials) and converted to scaled scores. An overall standardized scaled score based on a scale ranging from 1 (*impaired*) to 10 (*very superior*) was used for analysis. (2) The Color Trails Test *(CTT)* (D'Elia, Satz, Uchiyama, & White, 1996) was designed to measure shifting. CTT Part 2 showed numbers from 1 – 25 twice (each sequence has either a yellow or pink background) and required the participant to connect numbers in numerical order alternating between pink and yellow circles, disregarding the numbers in circles of the alternate color. A reverse coded latency score was used for analysis to be commensurate with all other tasks, thus higher scores indicate better performance.

*Inhibition.* (1) The Hayling Sentence Completion test was developed to index inhibition by measuring initiation speed and response suppression. It consists of two sections of 15 sentences, each missing the last word. Section 1 requires the participant to quickly and correctly complete the sentence, and measures response speed. Section 2 requires completing the sentence quickly with an unconnected word, and measures response suppression. Response latencies on both sections and errors on Section 2 are used to derive an overall scaled score for each participant on a scale ranging from 1 (*impaired*) to 10 (*very superior*). (2) The Stroop test measures inhibitory processes by requiring the respondent to name the color of the ink a word is printed in and supress the automatic response of reading the word itself (Taylor, Kornblum, Lauber, Minoshima, & Koeppe, 1997). The performance score is the interference index and reflects slowing in response to interference in Part C ([Part C time – Part A time]/Part A time). The interference index was reverse coded for the analyses to be commensurate with the other executive functioning tasks; therefore higher scores indicate better performance.

**EF Latent Model and Invariance Verification**. We used a systematic procedure to verify the psychometric characteristics of an EF latent variable**.** First, confirmatory factor analysis was performed to verify that a single latent EF factor fit this particular sub-sample of participants (see Supplementary Table 3 for goodness of fit information). Second, longitudinal measurement invariance was tested using (a) configural invariance, for which the same indicator variables load onto the latent variable to determine if the same EF measures represent the latent variable at each wave of data collection, (b) metric invariance, for which factor loadings are constrained to be equal for each latent variable indicating that each latent variable was measuring the same construct, and (c) scalar invariance, for which indicator intercepts are constrained to be equal allowing mean differences to be evident at the latent mean level. Third, EF factor scores were estimated in MPlus and used in all subsequent growth models.

Model fit for confirmatory factor analyses was determined using standard fit indices: (a) χ2 for which a good fit would produce a non-significant test (p > .05), indicating the data are not significantly different than the model estimates, (b) comparative fit index (CFI) for which ≥ .95 was judged a good fit and between .90 and .94 was judged an adequate fit, (c) root mean square error of approximation (RMSEA), for which ≤ .05 would be judged good and between .06 and.08 would be judged adequate, and (d) standardized root-mean-square residual (SRMR) for which good fit is judged by a value of ≤ .08 (Little, 2013). For latent growth modeling, we are using individually varying times of observation as parameters. Therefore, the traditional SEM model fit indices (i.e., Chi-square, RMSEA, CFI, etc) are not available. Instead, the loglikelihood (LL) and indices of relative fit, the Akaike (AIC) and Bayesian information criteria (BIC), are provided. The LL is the measure of the magnitude of the log-likelihood function for the particular combination of parameter estimates and observed data (Singer & Willett, 2003). It contains all parameters (sample data and the unknown parameters), and smaller absolute values mean a better model fit (Singer & Willett, 2003). To compare the nested growth models, the Deviance statistic (*D*) is used. The deviance statistic is = -2 (log-likelihood of the current model – log-likelihood of the saturated model; -2LL; Singer & Willett, 2003). For models that are hierarchically nested and use the same sample, the difference in deviance for competing models is computed and compared. The difference between the full (parameters free to vary) model and the reduced (constrained parameters) model has a χ2 distribution with degrees of freedom which is the number of constraints imposed (Singer & Willett, 2003). In the present analyses, the Δ*D* value was compared to a critical value with the appropriate degrees of freedom; a significant deviance statistic indicates a better model fit than the previous model.

We used overall model fit and Δχ2 tests to confirm measurement invariance (results presented in Thibeau et al., 2017), including the sequence from (a) configural invariance (χ2 = 33.68 *df* = 35, *p* = .53), (b) metric invariance (Δχ2 = 8.149, Δ*df* = 6, *p* = .23, ΔCFI = .001), and (c) scalar invariance (Δχ2 =152.37, Δ*df* = 8, *p* <.001, ΔCFI = .08). The significant effect for the latter indicated this criterion was not met, and thus we proceeded to test a model with partial scalar invariance (Δχ2 = 21.341 Δ*df* = 4, *p* <.001, ΔCFI = .01). Despite a significant decrease in model fit, we retained the partial scalar model with intercepts constrained to be equal across time for Stroop and Brixton, given the observation of a larger pattern of good fit indices and the acceptable ΔCFI (i.e., RMSEA = .027, CFI = .99, SRMR = .06, ΔCFI = .01). Two indicator variables (i.e., Hayling, Color Trails) exhibited mean differences outside of the latent variable. The invariance testing results indicated that the EF model measured the same construct over time, the same indicator variables marked EF at each wave, and partial scalar invariance allowed us to compare latent variable means (Kline, 2011).

#### Mobility and EPA as time-invariant and time-varying predictors. In the initial analyses of everyday activity and mobility predictions, we tested each variable in both potential prediction roles as time-invariant (level or baseline) and time-varying (change) predictors of cognitive change. The results were as follows. First we established the functional form of change (i.e., growth models) for EPA and mobility. The preferred growth model for EPA was a random intercept, fixed slope model. This model indicated that older adults varied in the level of EPA (b = .156, p < .001) and experienced significant decline in EPA (M = -.015, p < .003), but without individual differences in rate of decline in EPA (p > .05). The preferred growth model for mobility was a random intercept, random slope model, indicating that older adults varied in the level of mobility (b = .557, p < .001), experienced significant mobility decline (M = -.061, p < .001), and there was significant individual variability in rate of mobility decline (b = .002, p = .001).

#### To test the time-varying and time-invariant predictors, we first tested the EPA and mobility growth models in parallel process with the EF growth model. Time-varying EPA did not predict EF performance (b = -.002, p >.05) nor EF change (b = .529, p >.05). For mobility, level of mobility at predicted level of EF performance (b = .798, p =.003), but not change (b = .010, p =.81). In addition, change in mobility did not predict change in EF (b = .31, p > .70). We then tested mobility and EPA as time-invariant predictors (results presented in Thibeau et al., 2017 ), and with significant results we examined moderation of these relationships.

***Age as the metric of change.*** When multi-cohort sampling has occurred, an accelerated longitudinal design may be used. This design focuses on individualized trajectories across age (see Supplementary Figure S1 for executive function latent variable trajectories). In the present study, each individual’s EF score is related to that individual’s age at that point for both the intercept and the slope, and therefore there is no need to covary for age, as it is taken into account at each timepoint within the model. Notably, in our models, we are using individually varying timepoints, through the use of TSCORES (for age) in Mplus. This model adjusts for the within-person residual variance associated with the TSCORE variable (see Mplus user’s guide 6.12 and 9.16 to see how the TSCORE option considers age into the model (Muthén & Muthén, 1998 - 2012)). Additionally, we analyze these effects using a continuous variable, however for ease of visual interpretation (i.e., to represent the effect graphically) we use three groupings for the predictor variable (Little, 2013).

***Missing data and attrition.*** In the VLS we have adopted a specific protocol for dealing with missing data. Specifically, we (a) examine the source of the missing data, and then (b) use advanced statistical methods to account for missingness. First, Little’s tests of Missing Completely at Random within each sample were not significant (Sample 1: χ2 = 40.19, df = 662, *p* = 1.00; χ2 = 509.34, df = 4636, *p* = 1.00; Sample 2: χ2 = 727.02, df = 1026, *p* = 1.00; Sample 3: χ2 = 509.34, df = 4636, *p* = 1.00). This means the missingness within this data was completely at random. Therefore, we proceed with Maximum Likelihood based approaches. Second, we use a structural equation modeling (SEM) approach, which allows for the inclusion of data from all participants, even those with just one wave of data. Using robust maximum likelihood estimation methods based on all available information from every variable included in the covariance matrix, SEM estimates values for all waves (Little, 2013). Notably, within statistical growth curve models, attrition does not translate to subjects lost at all waves; there is no need for “listwise deletion” in these models, as all participants providing data at one, two, or all three waves are included in the analyses. Second, we account for missing data using multiple imputation. Multiple imputation refers to contemporary estimation of missing at random data (Enders, 2011; Little, 2013), and involves several steps. Statistical methodologists recommend 20 or more imputations as standard practice (Enders, 2011; Graham, Olchowski, & Gilreath, 2007), in the VLS, our protocol includes 50 imputations. Therefore, 50 datasets are generated each with a unique set of plausible replacement data. Then the analysis is conducted using each of the 50 generated datasets. Finally, the parameter estimates and standard errors are pooled into a single set of results.

# Results

## Foundational analyses: Latent growth modeling for EF; independent effects of EPA and mobility on the EF growth model.

The preliminary analyses confirmed the necessary and foundational effects pertaining to our basic growth models. Briefly, first we verified an EF latent varible and an EF growth model over a 40-year period. As expected, older adults varied significantly in level of EF performance (*b* = 1.05, *p* < 0.001), experienced significant decline in EF performance (*M* = -0.016, *p* <0.001), and had variability in rate of decline (*b* = 0.003, *p* <0 .001; see Supplementary Table 1 for full EF growth model results, Supplementary Material Figure 1 for growth model, and Supplementary Material Figure 2 for growth model trajectories across the 40 year band of aging; Thibeau et al., 2017).

Second, we examined activity and mobility as time-invariant (level or baseline) and time-varying (change) predictors of cognitive change. The results were as follows. The preferred growth model for EPA was a random intercept, fixed slope model. This model indicated that older adults varied in the level of EPA (*b* = .156, *p* < .001), and experienced significant decline in EPA (*M* = -.015, *p* < .003), but without individual differences in rate of decline in EPA (*p* > .05). The preferred growth model for mobility was a random intercept, random slope model, indicating that older adults varied in the level of mobility (*b* = .557, *p* < .001), experienced significant mobility decline (*M* = -.061, *p* < .001), and there was significant individual variability in rate of mobility decline (*b* = .002, *p* = .001; see Supplementary Material Table 4 for EPA and mobility growth model results).

Third, we tested the EPA and mobility growth models in parallel process with the EF growth model. Time-varying EPA did not predict EF performance (*b* = -.002, *p* >.05) nor EF change (*b* = .529, *p* >.05). For mobility, level of mobility at predicted level of EF performance (*b* = .798, *p* =.003), but not change (*b* = .010, *p* =.81). In addition, change in mobility did not predict change in EF (*b* = .31, *p* > .70).

Fourth, we tested mobility and EPA and time-invariant predictors (results presented in Thibeau et al., 2017). Briefly, results indicated that baseline EPA predicted both EF performance and change (Figure S1a), and baseline mobility predicted both EF performance and decline (Figure S1b). These results established the foundation for the specific goals of this paper.

## **Research Goal 1: Moderation of the EPA-EF and mobility-EF relationships by Sex**

*Sex moderation of the mobility-EF relationship.* As the pattern of results was similar for both males and females, we examined this moderation further by comparing models in which one parameter at a time was held to be equal across groups. When the intercepts and slopes were separately constrained to be equal across groups, the constrained model fit the data worse than the unconstrained model (*D* = 154.0, Δ*df* = 6, *p* < .001; *D* = 133.78, Δ*df* = 6, *p* < .001, respectively). Thisprovided evidence that there was a differential effect of baseline mobility on both EF performance and change between males and females. Specifically, males with lower baseline mobility experienced lower EF performance (*M* = -0.597) and steeper decline (*M* = -0.078) than did females with lower baseline mobility (*M* -0.275, *M* = -0.044, respectively).

**References**

Abner, E. L., Nelson, P. T., Kryscio, R. J., Schmitt, F. A., Fardo, D. W., Woltjer, R. L., . . . Xiong, C. (2016). Diabetes is associated with cerebrovascular but not Alzheimer's disease neuropathology. *Alzheimer's & Dementia, 12*(8), 882-889.

Aronow, W. S. (2017). Hypertension and cognitive impairment. *Annals of translational medicine, 5*(12).

Bielak, A. A., Mansueti, L., Strauss, E., & Dixon, R. A. (2006). Performance on the Hayling and Brixton tests in older adults: Norms and correlates. *Arch Clin Neuropsychol, 21*(2), 141-149. doi:10.1016/j.acn.2005.08.006

Brach, J. S., Talkowski, J. B., Strotmeyer, E. S., & Newman, A. B. (2008). Diabetes mellitus and gait dysfunction: possible explanatory factors. *Physical therapy, 88*(11), 1365-1374.

Burgess, P. W., & Shallice, T. (1997). *The Hayling and Brixton tests*. Thurston, Suffolk: Thames Valley Test Company.

D'Elia, L. F., Satz, P., Uchiyama, C. L., & White, T. (1996). *Color Trails Test: Professional Manual*. Odessa, FL: Psychological Assessment Resources.

de Frias, C. M., & Dixon, R. A. (2014). Lifestyle engagement affects cognitive status differences and trajectories on executive functions in older adults. *Arch Clin Neuropsychol, 29*(1), 16-25. doi:10.1093/arclin/act089

de Frias, C. M., Dixon, R. A., & Strauss, E. (2006). Structure of four executive functioning tests in healthy older adults. *Neuropsychology, 20*(2), 206-214. doi:10.1037/0894-4105.20.2.206.

de Frias, C. M., Dixon, R. A., & Strauss, E. (2009). Characterizing executive functioning in older special populations: From cognitively elite to cognitively impaired. *Neuropsychology, 23*(6), 778-791. doi:10.1037/a0016743

Dixon, R. A., & de Frias, C. M. (2004). The Victoria Longitudinal Study: From characterizing cognitive aging to illustrating changes in memory compensation. *Aging Neuropsychol Cog, 11*(2-3), 346-376. doi:10.1080/13825580490511161

Enders, C. K. (2011). Analyzing longitudinal data with missing values. *Rehabilitation Psychology, 56*(4), 267-288. doi:org/10.1037

Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science, 8*(3), 206-213.

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (Vol. 3rd). New York, NY: Guilford Press.

Little, T. D. (2013). *Longitudinal structural equation modeling*. New York, NY: Guilford Press.

McFall, G. P., Wiebe, S. A., Vergote, D., Jhamandas, J., Westaway, D., & Dixon, R. A. (2014). IDE (rs6583817) polymorphism and pulse pressure are independently and interactively associated with level and change in executive function in older adults. *Psychol Aging, 29*(2), 418-430. doi:org/10.1037/a0034656

McFall, G. P., Wiebe, S. A., Vergote, D., Westaway, D., Jhamandas, J., & Dixon, R. A. (2013). IDE (rs6583817) polymorphism and type 2 diabetes differentially modify executive function in older adults. *Neurobiology of aging, 34*(9), 2208-2216. doi:10.1016/j.neurobiolaging.2013.03.010

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive psychology, 41*(1), 49-100.

Moheet, A., Mangia, S., & Seaquist, E. R. (2015). Impact of diabetes on cognitive function and brain structure. *Annals of the New York Academy of Sciences, 1353*(1), 60-71.

Muthén, L. K., & Muthén, B. O. (1998 - 2012). *Mplus User's Guide* (Vol. Seventh Edition). Los Angeles, CA: Muthén & Muthén.

Nandipati, S., Luo, X., Schimming, C., Grossman, H. T., & Sano, M. (2012). Cognition in Non-Demented Diabetic Older Adults. *Current aging science, 5*(2), 131-135.

Novak, V., & Hajjar, I. (2010). The relationship between blood pressure and cognitive function. *Nature Reviews Cardiology, 7*, 686. doi:10.1038/nrcardio.2010.161

Rosano, C., Longstreth, W. T., Boudreau, R., Taylor, C. A., Du, Y., Kuller, L. H., & Newman, A. B. (2011). High blood pressure accelerates gait slowing in well-functioning older adults over 18-years of follow-up. *Journal of the American Geriatrics Society, 59*(3), 390-397. doi:10.1111/j.1532-5415.2010.03282.x

Sapkota, S., Wiebe, S. A., Small, B. J., & Dixon, R. A. (2015). Apolipoprotein E and Clusterin can magnify effects of personality vulnerability on declarative memory performance in non-demented older adults. *Int J Geriatr Psychiatry*. doi:10.1002/gps.4355

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York, NY: Oxford University Press; US.

Taylor, S. F., Kornblum, S., Lauber, E. J., Minoshima, S., & Koeppe, R. A. (1997). Isolation of specific interference processing in the Stroop task: PET activation studies. *Neuroimage, 6*(2), 81-92. doi:10.1006/nimg.1997.0285

Thibeau, S., McFall, G. P., Wiebe, S. A., Camicioli, R., & Dixon, R. A. (2017). Alzheimer's disease biomarkers interactively influence physical activity, mobility, and cognition associations in a non-demented population. *Journal of Alzheimers Disease, 60*(1), 69-86. doi:10.3233/JAD-170130.

Yaffe, K., Falvey, C., Hamilton, N., Schwartz, A. V., Simonsick, E. M., Satterfield, S., . . . Harris, T. (2012). Diabetes, glucose control and 9 year cognitive decline among non-demented older adults. *Archives of neurology, 69*(9), 1170-1175. doi:10.1001/archneurol.2012.1117

**Table Titles**

Supplementary Table 1

Descriptive Statistics for Sample by *APOE* Genotype

Supplementary Table 2

Confirmatory Factor Analyses and Latent Growth Goodness of Fit Indexes for Executive Function Models

Supplementary Table 3

Confirmatory Factor Analysis for EF latent variable

Supplementary Table 4

Latent Growth Goodness of Fit Indexes for Everyday Physical Activity and mobility models

Supplementary Table 5

Covariate adjusted results

**Figure Captions**

*Supplementary Figure 1.* Executive Function (EF) latent growth model, with individually varying times of observation, and a random intercept and random slope (final growth model; *D* = 1058.46, Δdf = 2, *p* <.001). Curved double-headed lines represent variance parameters, straight single-headed lines represent regression parameters

*Supplementary Figure 2.* Individual Executive Function (EF) trajectories across a 40-year band of aging with a group mean trajectory line.

*Supplementary Figure 3.* Predicted growth curve model of Executive Function (EF) with continuous Everyday Physical Activity (EPA) or mobility as a predictor. The three categories of EPA and mobility are depicted for convenience. Age in years was the metric of change. The age variable was centered at 75 years. Baseline level of EPA predicted both EF performance (*b* = 0.340, *p* < .001) and 9-year change (*b* = 0.016, *p* = .002; Figure S1a). Baseline level of mobility predicted EF performance (*b* = 0.412, *p* < .001) and 9-year decline (*b* = 0.025, *p* < .001; Figure S1b). Reproduced from Thibeau et al., 2017.

*Supplementary Figure 4*. Predicted growth curve model of Executive Function (EF) using EPA as a predictor and moderated by sex. The three categories of EPA are depicted for convenience. Age in actual years was the metric of change. The age variable was centered at age 75.

*Supplementary Figure 5.* Predicted growth curve of Executive Function (EF) using mobility as a predictor and moderated by sex. The three categories of mobility are depicted for convenience. Age in actual years was the metric of change. The age variable was centered at age 75.