Gradual Change, Homeostasis, and Punctuated Equilibrium:

Reconsidering Patterns of Health in Later-Life

Michal Engelman[1] and Heide Jackson[1,2]

[1] Center for Demography of Health and Aging, University of Wisconsin

[2]Maryland Population Center, Univeristy of Maryland

Methodological Appendix

This appendix provides additional information about the five analytic methods used to generate the empirical results in this paper. While our key finding is that stability is more common than change in the Health and Retirement survey data, the data also record substantial item missingness, wave non-response, individual dropout, and mortality. The various forms of missing data affect the sample composition over time, and introduce substantial uncertainty about the congruence between the health changes experienced by respondents, what is documented in the survey data, and the methods used to model longitudinal health patterns. We begin by describing patterns of missing data in the HRS before moving to more detailed considerations of each method. For the latent growth curve, latent class growth analysis, and multistate model, we provide key model fit statistics. We also discuss supplemental analyses for the various methods, including robustness and sensitivity checks. Finally, we describe the implications of missing data for inferences from each of the five methods.

Prevalence of Missing Data in the Health and Retirement Study

While no health change is the most common outcome for HRS respondents who remained in the study, **Table A1** shows that 52% of the sample is lost to attrition or mortality before the end of the follow up period. Those who dropped out or died were disproportionately likely to be dropped from the LGC, LCGA, and descriptive analyses. 23% of respondents who dropped out and 18% of respondents who died are excluded from these analyses because they contribute one or fewer reports of functional limitations.[1] Respondents who died and contributed just one round of data had substantially more functional limitations when observed compared to cases who contributed

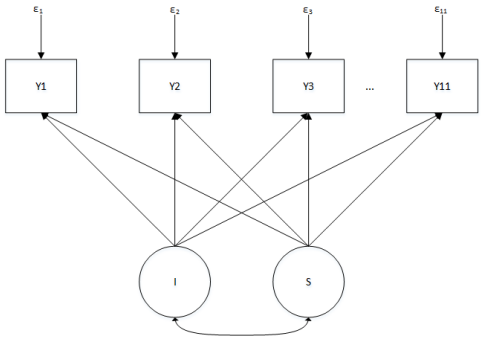
[1]Respondents would be dropped if they remained in sample but contributed only to the last survey round.

to multiple survey rounds before leaving the sample. Both prior studies (Jackson et al., 2019; Zajacova and Burgard, 2013) and our own supplementary analyses (available upon request) show that members of the HRS cohort who contribute multiple waves of data differ considerably from those who are excluded in analyses of longitudinal change.

As we discuss below, within the gradualist analytic framework there are advanced techniques to account for missing data in longitudinal analyses of health, including the use of full-information maximum likelihood and inverse probability weighting. However, as we explain below, while these methods enable a more complete use of available data, they do not correct for the non-random selection processes at play. Subject dropout and mortality are very common events that alter the sample composition and therefore influence our inferences about the sample's health status and health changes throughout the longitudinal study follow-up period.[2] Together, our findings suggest that the analytic sample is selected to be healthier than the original HRS sample.

Growth Curve Modeling

Formal Specification. The latent growth curve results shown in the main text were estimated in MPlus 7.11 based on the following model:



This univariate latent growth curve is shaped by functional limitations measured at 11 time points (Y1, Y2,...,Y11) and described parametrically by an intercept (I) and a linear slope (S).

The model can also be summarized by the following equation:

$$Y_{it} = \eta_{0i} + \eta_{1i} x_{it} + \varepsilon_i \tag{1}$$

[2] Due to limitations of the data, we cannot precisely determine the timing of subject dropout and mortality. Some respondents may have dropped out of the sample due to death, while in other cases death may have occurred after a respondent left the sample

where Y_{it} is the count of functional limitations for individual i at time t; η_{0i} is the intercept of the functional limitation trajectory for individual i; η_{1i} is the rate of change (slope) in the number of functional limitations for individual i across different time periods (x_t); and ε_{it} represents random error in functional limitations.

The person-specific growth curve slope and intercept parameters are allowed to vary randomly and are estimated as dependent variables by the following person-level models:

$$\eta_{0i} = \alpha_{00} + \alpha_{01} w_i + \zeta_{0i} \tag{2}$$

$$\eta_{1i} = \alpha_{10} + \alpha_{11} w_i + \zeta_{1i} \tag{3}$$

where the α parameters represent the slope and intercept of each growth parameter and w are time invariant covariates. ζ_{0i} and ζ_{1i} are random effects with mean of zero.

Fit Statistics. **Table A2** shows the fit statistics for latent growth curve models with linear and quadratic slopes. The lower AIC, BIC, and RMSEA statistics, along with the higher CFI and TLI statistics are consistent in recommending the latent growth curve model with a quadratic slope shown in Table 1 and Figure 2 in the main text as the best fitting model.

Robustness and Sensitivity Checks. Although not shown here, it is possible within the latent growth curve model to estimate individual trajectories which may deviate from the average population trajectory. One could extend this further by interacting individual characteristics with the intercept and slope terms, to identify those who report no health change or show evidence of a recovery. Nonetheless, due to the parametric constraints in model fitting, such individual trajectories would still appear smooth and represent change as gradual. We do not present this analysis in the main text or in this appendix because we believe the resulting estimates are likely biased due to the presence of non-random missing data.

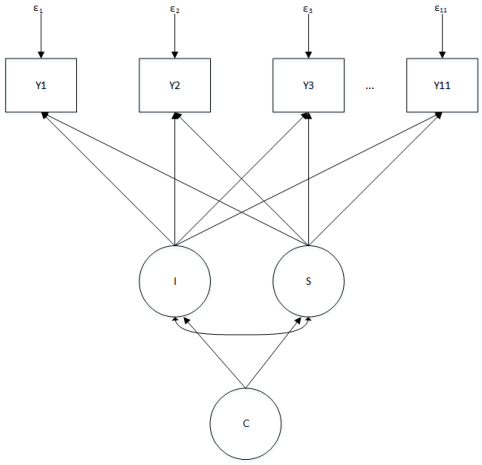
The Impact of Missing Data on Model Inferences. Our paper adopts the standard approach to estimating growth curve models, full information maximum likelihood (FIML). The oft-cited advantage of FIML estimation is that it allows respondents to contribute information until the time of dropout. However, FIML assumes that data is missing at random after conditioning on

observed factors. In longitudinal studies of the health of older adults, however, we often have strong evidence suggesting that missingness is not random. Rather, health shocks and health declines in the two years between survey rounds may directly influence the likelihood of respondents having missing data or being lost to follow up.

Although we cannot conclusively demonstrate here that data is not missing at random, prior research (Zajacova and Burgard, 2013; Jackson et al., 2019) suggests that the HRS has significant non-randomly missing data. Because individuals who dropout likely have poorer health outcomes including the accumulation of more functional limitations, latent growth curve trajectories likely underestimate the burden of functional limitations in the population and the true shape of the average trajectory is not known.

Latent Class Growth Analysis

Formal Specification The latent class growth class growth models (LCGA) shown in the main text were estimated in MPlus 7.11 based on the model depicted here:



This LCGA model is similar to the latent growth curve model specified above but includes the addition of a categorical variable (c) which indicates latent class:

$$Y_{it}^{c} = \eta_0^c + \eta_1^c x_{it} + \varepsilon_{it} \tag{4}$$

Y_{it}^{c} represents the functional status of individual i at time t given membership in latent class c. The η parameters represent the coefficients associated with the intercept and rate of change in functional status for individuals in class c. ε_{it} is a disturbance term assumed to be normally distributed with mean zero and constant variance. As in the latent growth curve model described above, each latent

class has a class-specific intercept and slope predicted using equations equivalent to (2) and (3) above.

Fit Statistics. Table A3 shows the sample adjusted BIC for fitted LCGA models with varying numbers of classes and slope types. The best-fitting model selected and displayed in the main text is the four class quadratic, which has the smallest sample adjusted BIC and a significant Lo-Mendell-Rubin likelihood ratio test. The sample adjusted BIC is not shown for the five class quadratic and six class models because these models did not converge. The sample adjusted AIC also suggested the four class quadratic model provides the best fit to the data (results available upon request). Although these statistics generally agree in recommending the four class quadratic model as the preferred fit, prior studies caution against relying solely on fit statistics in selecting models (Tein et al., 2013; Tofighi and Enders, 2008).

Robustness and Sensitivity Checks. Because of the uncertainty in the preferred number of latent classes, we examined the trajectories for all models that achieved convergence (results available upon request). Across models, trajectories are smooth and there is one latent class that displays a slow, steady progression of limitations and is estimated to contain a majority of the sample (see Class A of Figure 3 in the main text). Additionally, at least one of the other fitted classes has a high level of impairment that remains fairly constant over time, similar to Class C. However, we caution against the tendency to reify any latent class due to concerns about non-randomly missing data.

The Impact of Missing Data on Model Inferences. In sensitivity analyses, we find that as in the case of the latent growth curves, LCGAs modeled with a FIML approach underestimate functional limitations for the sub-sample that dies over the course of follow up. Furthermore, the exact shape of specified latent classes is sensitive to assumptions made about cases with missing data (results available upon request). The likely consequences of the violation of the FIML assumption of random missingness are (1) an underestimation of the prevalence and progression of functional limitations and (2) substantial uncertainty regarding the shape of estimated latent classes (Jackson et al., 2019).

Furthermore, there is considerable uncertainly in the assignment of individuals into latent classes. As we show via the descriptive, multistate, and sequence analyses, the LCGA does not

seem to accurately represent observed individual trajectories. Moreover, for individuals who drop out of the sample, there can be substantial class misclassification (see also Warren et al. (2015)).

Descriptive Analysis

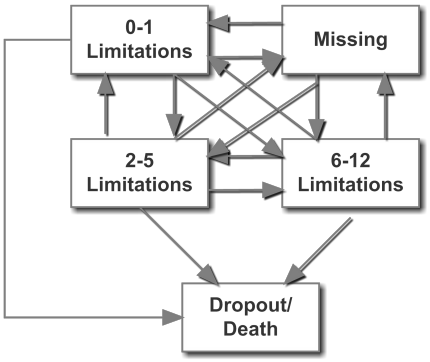
Supplemental Analysis. To supplement the findings reported in the main text, we calculated how many individuals (of those who remain in the sample over time) experience increasing functional limitations across successive survey rounds. The key findings in **Table A4** is that no individual experiences the successive, gradual progression of functional limitations implied by the LGC and LCGA analyses. All descriptive analyses were conducted using R.

The Impact of Missing Data on Model Inferences. Notably, descriptive analyses are also influenced by the presence of missing data. In the main analysis, individuals are included in Table 2 only if they contribute health information to at least two survey rounds. This reduces the representativeness of the sample and deflates the proportion with constant health over time, rendering our estimates of stability conservative. However, the extent of item-specific and wave missingness is not captured in the table. We might expect that individuals who contribute information to more rounds of data would give us more opportunities to observe a health change while cases with missing information may be less likely to have a health change captured in the survey data, whether or not they actually experienced a health change.

Multistate Analysis

Formal Specification

As shown in the main text, the multi-state model we calculated using Stata contains 5 mutually exclusive and exhaustive states:



The corresponding transition matrix is:

$$Q = \begin{pmatrix} q_{11} & q_{12} & q_{13} & q_{14} & 0 \\ q_{21} & q_{22} & q_{23} & q_{24} & 0 \\ q_{31} & q_{32} & q_{33} & q_{34} & 0 \\ q_{41} & q_{42} & q_{43} & q_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{5}$$

Transitions are governed by a Markov process with continuous time where state occupancy is observed a finite number of times. The likelihood function for this model is calculated from the transition probability matrix of:

$$P(t) = Exp(tQ) \tag{6}$$

and a likelihood function that is the product of transition probabilities between states $S(t_j)$ at successive times (t_j):

$$L(Q) = \prod_j P_{S(tj),S(tj+1)}(t_{j+1} - t_j) \tag{7}$$

Fit Statistics. A comparison of fit statistics for alternative specifications of the multistate model is shown in **Table A5**. Results confirm that the model shown in the main text has the lowest AIC, indicating the best fit to the data. Models with more than three piecewise time breaks did not converge. **Table A6** shows that the percentages of people in each state as predicted by the best-fitting multistate model track well with those observed in each state at each time period.

Robustness and Sensitivity Checks. A key limitation of the multistate approach is that it collapses functional limitations from 13 to 3 categories. In the main analysis, the three states are (1) 0 to 1 limitations, (2) 2 to 5 limitations, and (3) 6 to 12 limitations. Because most respondents have only a few limitations, this model will necessarily miss health changes that occur within the three states. To address this concern, we conducted sensitivity analyses that varied the cutpoints for the health states, and found that results were qualitatively similar to those shown in the main text, supporting the robustness of the conclusions. Results from one particularly interesting sensitivity check are shown in **Table A7**. This model defines the three health states as having 0 limitations (state 1), 1 to 4 limitations (state 2), and 5 to 12 limitations (state 3). By keeping individuals with

no health limitations in a separate state, we are able to effectively estimate the transition from having no to any limitation. Again, the finding is that remaining in the same state is the most common outcome between any two survey rounds.

The Impact of Missing Data on Model Inferences. Unlike the LGC, LCGA, and descriptive analyses, the multistate model we specified explicitly defines temporary missingness as a state. It shows the probability of transitioning between the observed health states and the temporary missing state as well as the probability of transitioning into a permanent dropout or mortality state. Table 3 in the main text shows that the probability of transitioning into a temporary missing state is similar across observed health states. This finding suggests that there is not strong health selection driving item or temporary missingness. However, people with more functional limitations are more likely to dropout of the sample or die in the course of follow up. For example, at survey round 2, people with one or no limitations have a 4% chance of leaving the sample at the next survey wave while people with 5-12 limitations have a 10% chance of leaving the sample. These findings suggest that our multistate analysis – like its LGC and LCGA counterparts – may disproportionately under represent individuals with severe health limitations.

Sequence Analysis

Sequence analyses were conducted in Stata using SQ-Ados.

The Impact of Missing Data on Model Inferences. Like the multistate model, the sequence analysis explicitly models missingness as a possible outcome across survey rounds. The most striking finding from Figure 4 in the main text is that the second most common trajectory, characterizing 219 individuals in the sample, is dropout before survey round 2, and the third most common trajectory, characterizing 171 persons, is death before survey round 2. Most of the most common sequences include some periods when an individual had missing data or had dropped out of the sample. Findings demonstrate that the prevailing experience of individuals is one of no change or consistently missing measures of functional limitation.

Incorporating Uncertainty across Model Specifications

To this point, this paper has been largely silent on the impact of measurement error on inferences across model specifications. Below, we briefly describe sensitivity analyses that could explicitly address stochastic error in each model specification.

In latent growth curve models, there is an error term which allows for individual values to vary around the average estimated trajectory. It is possible to leverage this error term in predictive models to generate a confidence interval around the fitted average trajectory. In the latent class growth analyses, an equivalent uncertainty can be quantified for individual predicted values. Since individuals are also classified into particular latent classes with some error, it is also possible to rerun the classification model to see how an individual's predicted latent class varies across model draws (Nagin, 2005). While both of these exercises may be conducted to characterize the uncertainty in predictions from latent growth curve and latent class models, the average fitted trajectories will always emphasize the gradualist parametric assumptions built into these regression models.

Multi-state and sequence analyses are also susceptible to individual classification error and uncertainty in population estimates. A multi-state model estimates a matrix of transition probabilities for a population with an observed initial distribution of states and distributions of states observed at some finite number of follow-up time points. A simulation could extend this model by generating some number of hypothetical cohorts with an initial distribution of states equal to that observed in the actual population. The estimated transition matrix can then be applied to each hypothetical cohort for a given follow-up period, yielding a set of predicted values for each hypothetical cohort at each observation time. These values can be averaged or examined at the 2.5, 50, or 97.5 percentile of draws to characterize the expected variation in state occupancy at each follow-up point.

Although not commonly done, uncertainty in the sequence analysis could also be quantified. One strategy might combine the multi-state models with the sequence analysis: after estimating the expected state distribution of a set of hypothetical cohorts, it would be possible to run a sequence analysis on each cohort. Sequences at the 2.5, 50, or 97.5 percentiles of the simulated cohorts can then be compared to bound the estimates from the observed HRS sample. Because the sequence analysis depends on the details of repeated survey measures, it may also be influenced by reporting error, though it is not possible to empirically distinguish this from other sources of stochasticity

captured by the sensitivity analysis.

The procedures described above represent methods of quantifying the sensitivity of estimates in the main text to measurement error, but they do not change the underlying assumptions implied by each of the methods described. It is important to note that analyses of population data cannot separate measurement error due to inaccurate or incomplete reporting from variation due to pure chance or a set of highly contingent events that cannot be fully documented in quantitative data. On the aggregate level, these distinctions matter little as inferences are based on average trajectories, which are well-represented by gradualist models. The many contributions to the messy heterogeneity observed in individual histories are what the punctuated equilibrium perspective calls attention to.

Table A1: Summary of Missing Data in the Health and Retirement Study

	In-Sample N=4912		Dropout N=1781		Died N=3505	
	Percent	Mean at First Round Observed	Percent	Mean at First Round Observed	Percent	Mean at First Round Observed
Chronic Conditions						
Report Chronic Conditions in at Least Two Rounds	1.00	0.84	0.87	0.83	0.91	1.49
Report Chronic Conditions in One Round	0.00	3.00	0.13	1.11	0.09	2.04
Do Not Report Chronic Conditions	0.00	–	0.00	–	0.00	–
Functional Limitations						
Report Functional Limitations in at Least Two Rounds	1.00	1.52	0.75	1.37	0.83	2.62
Report Functional Limitations in One Round	0.00	6.00	0.13	1.37	0.10	3.62
Do Not Report Functional Limitations	0.00	–	0.10	–	0.08	–
Activities of Daily Living (ADLs)						
Report Limitation in ADLs in at Least Two Rounds	1.00	0.09	0.75	0.08	0.83	0.25
Report Limitation in ADLs in One Round	0.00	0.00	0.13	0.06	0.10	0.58
Do Not Report Limitation in ADLs	0.00	–	0.12	–	0.08	–

Table A2: Fit Statistics for Latent Growth Curve Models

Model	AIC	SABIC	RMSEA	CFI	TLI
Linear Latent Growth Curve	311461.707	311525.749	0.073	0.950	0.955
Quadratic Latent Growth Curve	309720.802	309800.855	0.051	0.977	0.978

Table A3: Fit Statistics for LCGA Models

Model	Sample Adjusted BIC
2 class linear	308003.334
2 class quadratic	306178.625
3 class linear	306664.58
3 class quadratic	304585.4
4 class linear	305882.043
4 class quadratic	303683.76
5 class linear	305273.165
5 class quadratic	–
6 class linear	–
6 class quadratic	–

Table A4: Prevalence of Gradually Increasing Functional Limitations

Successive Increase Through	Starting Sample Observed at Round 2 N=8,825
Round 3	2,487
Round 4	504
Round 5	77
Round 6	12
Round 7	5
Round 8	1
Round 9	0

Table A5: Fit Statistics for Alternative Multistate Model Specifications

Model	AIC
Transition Probabilities Constrained to be Constant Across Rounds	159407.1
Transition Probabilities Constrained to be Constant Between Rounds 2 to 5, 6 to 12	158736.4
Transition Probabilities Constrained to be Constant Between Rounds 2 to 5, 6 to 8, 9 to 12	158209.4

Table A6: Observed and Multistate Model Predicted Percent of Sample in Each Health State Across Survey Rounds

Round	Observed	Predicted	Observed	Predicted	Observed	Predicted	Observed	Predicted	Observed	Predicted
	0 to 1 Limitations		2 to 5 Limitations		6 to 12 Limitations		Missing		Dropout/Died	
2	54.34	54.34	27.56	27.56	9.03	9.03	9.08	9.08	0.00	0.00
3	51.51	50.01	24.82	26.73	10.70	10.55	8.36	7.80	4.61	4.91
4	47.54	46.34	25.15	25.75	10.75	11.19	7.20	7.00	9.37	9.72
5	43.70	43.20	24.06	24.67	10.97	11.33	7.36	6.44	13.91	14.36
6	39.11	40.46	25.90	23.54	11.08	11.19	5.17	6.01	18.74	18.81
7	35.40	36.00	26.18	24.84	11.41	11.32	4.54	4.92	22.47	22.92
8	31.63	32.82	25.66	24.57	11.86	11.38	4.36	4.30	26.49	26.93
9	29.18	30.34	25.19	23.70	11.60	11.26	3.20	3.90	30.83	30.80
10	24.51	26.88	23.46	22.40	12.83	11.90	2.26	2.38	36.94	36.43
11	22.91	23.88	21.48	20.69	12.26	11.71	1.24	1.92	42.11	41.80
12	18.98	21.38	19.32	18.95	12.26	11.15	0.00	1.71	49.44	46.80

Table A7: Alternative Specification for Multistate Models

Rounds 2 to 5					
	0 Limitations	1 to 4 Limitations	5 to 12 Limitations	Missing	Dropout/Died
0 Limitations	0.64	0.25	0.03	0.03	0.04
1 to 4 Limitations	0.21	0.60	0.12	0.03	0.05
5 to 12 Limitations	0.04	0.21	0.64	0.03	0.09
Missing	0.19	0.19	0.09	0.52	0.01
Dropout/Died	0.00	0.00	0.00	0.00	1.00
Rounds 6 to 8					
	0 Limitations	1 to 4 Limitations	5 to 12 Limitations	Missing	Dropout/Died
0 Limitations	0.61	0.29	0.03	0.03	0.04
1 to 4 Limitations	0.15	0.65	0.13	0.02	0.04
5 to 12 Limitations	0.03	0.21	0.64	0.02	0.10
Missing	0.16	0.22	0.12	0.49	0.02
Dropout/Died	0.00	0.00	0.00	0.00	1.00
Rounds 9 to 12					
	0 Limitations	1 to 4 Limitations	5 to 12 Limitations	Missing	Dropout/Died
0 Limitations	0.58	0.30	0.04	0.01	0.06
1 to 4 Limitations	0.13	0.61	0.17	0.01	0.08
5 to 12 Limitations	0.02	0.17	0.64	0.01	0.17
Missing	0.18	0.29	0.23	0.25	0.05
Dropout/Died	0.00	0.00	0.00	0.00	1.00

References

H. Jackson, M. Engelman, and K. Bandeen-Roche. Robust respondents and lost limitations: The implications of nonrandom missingness for the estimation of health trajectories. *Journal of aging and health*, 31(4):685–708, 2019.

D. Nagin. *Group-based modeling of development*. Harvard University Press, 2005.

J.-Y. Tein, S. Coxe, and H. Cham. Statistical power to detect the correct number of classes in latent profile analysis. *Structural equation modeling: a multidisciplinary journal*, 20(4):640–657, 2013.

D. Tofighi and C. K. Enders. Identifying the correct number of classes in growth mixture models. *Advances in latent variable mixture models*, (Information Age Publishing, Inc):317–341, 2008.

J. R. Warren, L. Luo, A. Halpern-Manners, J. M. Raymo, and A. Palloni. Do different methods for modeling age-graded trajectories yield consistent and valid results? *American Journal of Sociology*, 120(6):1809–1856, 2015.

A. Zajacova and S. A. Burgard. Healthier, wealthier, and wiser: a demonstration of compositional changes in aging cohorts due to selective mortality. *Population research and policy review*, 32(3):311–324, 2013.