Rejoinder to "The Prevalence of Cognitive Impairment Is Not Increasing in the United States: A Critique of Hale et al. (2020)"

Jo Mhairi Hale,[a,b] *Daniel C. Schneider,*[b] *Jutta Gampe,*[b] *Neil K. Mehta,*[c] *and Mikko Myrskylä*[b,d]

We thank Lee et al.[1] for engaging with the article by Hale et al.[2] "Trends in the Risk of Cognitive Impairment in the United States, 1996–2014" ("Hale et al.[2]"). Both research teams agree with the original article's main point: when modeling trends in cognitive health using longitudinal data, practice effects (PEs) and nonrandom attrition must be addressed. The teams, however, reached opposite substantive conclusions.

Hale et al.[2] document that dementia trends are not declining. We reached this conclusion after a careful accounting of many potential biasing factors, including practice effects and selective mortality. Lee et al.[1] argue that Hale et al.[2] failed to account for selective attrition. Using a series of regressions, Lee et al.[1] argue that if practice effects (or panel conditioning) and attrition were jointly accounted, the results suggest a declining dementia trend.

This rejoinder has two sections. We first argue that the proposed alternative modeling strategy by Lee et al.[1] is problematic. We then acknowledge and discuss their conceptual contribution. We conclude that more refined estimation strategies, including using complementary data sources, may be necessary to precisely identify trends, an undertaking beyond the scope of this rejoinder.

The differences between the Hale et al.[2] and Lee et al.[1] model specifications are that Lee et al.[1]:

(1) uses a slightly different analytical sample,
(2) measures practice effects in a way we deem questionable,
(3) includes an interaction between current interview count and proxy interview marker,
(4) includes the total number of interviews ever taken, including future interviews, as a proxy for propensity-to-attrit.

Lee et al.[1] do not explain differences (1)–(2) above. They claim that (3) improves estimation of panel conditioning. They argue that (4) is the key difference that adjusts for attrition. We analyze the implications of each of these differences by sequentially implementing the changes. Thanks to the provision of replication code by Lee et al.,[1] we can exactly replicate their models.

The following paragraphs describe the models in the Table and are numbered according to the sequence of differences outlined above:

TABLE. Odds Ratios (Exponentiated Coefficients) and 95% Confidence Intervals From Logistic Regression Models for Dementia

	Baseline Model: Hale et al.					
			Model Modifications			
			Lee et al. IC	Lee et al. IC	Lee et al. IC	
					Lee et al. IC × Proxy	
	Hale et al. Model			Lee et al. IC × Proxy (= Lee et al. Model 2)	Lee et al. Total Interview (= Lee et al. Model 3)	Lee et al. Total Interview
	Hale et al. Sample		Lee et al. Sample			
Variable	(A)	(B)	(C)	(D)	(E)	(F)
Time trend (unit: 10 years)	1.29 (1.16, 1.44)	1.32 (1.20, 1.45)	1.22 (1.11, 1.35)	1.01 (0.87, 1.17)	0.71 (0.63, 0.81)	1.30 (1.19, 1.43)
Interview/test count (reference 1st)						
2nd	0.91 (0.83, 1.00)	0.90 (0.82, 1.00)	0.93 (0.84, 1.03)	0.98 (0.89, 1.09)	1.18 (1.07, 1.31)	0.94 (0.85, 1.03)
3–4	0.66 (0.59, 0.74)	0.65 (0.59, 0.73)	0.69 (0.62, 0.78)	0.77 (0.68, 0.87)	1.20 (1.07, 1.34)	0.72 (0.65, 0.80)
5–7	0.46 (0.40, 0.53)	0.45 (0.40, 0.52)	0.54 (0.46, 0.62)	0.65 (0.54, 0.78)	1.53 (1.32, 1.79)	0.58 (0.52, 0.66)
8+	0.33 (0.27, 0.40)	0.32 (0.27, 0.39)	0.41 (0.34, 0.49)	0.54 (0.43, 0.69)	1.85 (1.52, 2.26)	0.46 (0.38, 0.55)
Proxy × IC (reference 1st)						
2nd				1.24 (0.97, 1.59)	1.27 (0.98, 1.64)	
3–4				2.25 (1.86, 2.74)	2.31 (1.88, 2.83)	
5–7				3.05 (2.49, 3.74)	3.15 (2.53, 3.91)	
8+				4.79 (3.63, 6.32)	5.37 (4.04, 7.14)	
Total interviews (reference 1st)						
2nd					1.33 (0.96, 1.84)	1.45 (1.02, 2.08)
3–4					0.95 (0.70, 1.30)	1.16 (0.83, 1.62)
5–7					0.70 (0.51, 0.95)	1.15 (0.82, 1.62)
8+					0.36 (0.26, 0.49)	0.75 (0.53, 1.06)
Age (unit: 10 years)						
Slope	3.16 (2.70, 3.70)	3.18 (2.71, 3.73)	3.04 (2.59, 3.55)	2.70 (2.29, 3.18)	2.63 (2.22, 3.13)	3.52 (2.98, 4.17)
Quadratic	0.98 (0.96, 1.01)	0.98 (0.95, 1.01)	0.99 (0.96, 1.02)	1.00 (0.97, 1.03)	1.00 (0.97, 1.03)	0.96 (0.93, 0.99)
Race/ethnicity (reference White)						
Black	4.31 (3.78, 4.91)	4.25 (3.72, 4.85)	4.29 (3.76, 4.91)	4.36 (3.81, 4.99)	4.39 (3.82, 5.04)	4.23 (3.70, 4.84)
Latinx	3.18 (2.72, 3.73)	3.13 (2.67, 3.68)	3.20 (2.72, 3.76)	3.26 (2.77, 3.84)	3.34 (2.84, 3.93)	3.15 (2.69, 3.70)
Other	1.94 (1.13, 3.36)	1.87 (1.07, 3.27)	1.89 (1.08, 3.29)	1.94 (1.11, 3.38)	1.88 (1.07, 3.31)	1.81 (1.02, 3.19)
Men (reference women)	1.00 (0.92, 1.09)	0.98 (0.91, 1.06)	1.01 (0.94, 1.09)	1.02 (0.95, 1.10)	0.98 (0.91, 1.06)	0.96 (0.89, 1.04)
Proxy response (reference self-response)	10.69 (9.36, 12.20)	12.54 (10.91, 14.42)	13.93 (12.10, 16.02)	9.36 (7.62, 11.49)	7.98 (6.45, 9.87)	13.63 (11.88, 15.65)
Phone interview (reference face-to-face)	0.78 (0.72, 0.86)	0.77 (0.72, 0.82)	0.77 (0.72, 0.83)	0.76 (0.71, 0.81)	0.77 (0.72, 0.83)	0.79 (0.74, 0.84)
Proxy × men	0.37 (0.32, 0.43)	0.38 (0.33, 0.44)	0.37 (0.32, 0.43)	0.37 (0.32, 0.42)	0.39 (0.34, 0.45)	0.40 (0.34, 0.46)
Phone × men	0.95 (0.84, 1.08)					
Constant	0.004 (0.003, 0.005)	0.004 (0.003, 0.005)	0.004 (0.003, 0.005)	0.005 (0.004, 0.006)	0.007 (0.005, 0.01)	0.004 (0.002, 0.006)
N	179,236	180,159	180,159	180,159	180,159	180,159

Column (A) shows the coefficient estimates reported in Hale et al. Column (B) shows the same model applied to the data sample by Lee et al. It is the reference model for Columns C–F, as all these models use the same sample and differ only by the following: Column (C) uses IC by Lee et al. Column (D) adds an interaction of Lee et al. IC to proxy marker (Model 2 in Lee et al.). Column (E) further controls for the total number of interviews taken by a subject (Model 3 in Lee et al.). Finally, Column (F) applies variable definitions from the original Hale et al. specifications, but adds a control for the total number of interviews.

The original model of Hale et al. (Column A) has a slightly different specification for sex, proxy, and interview mode. Coefficients on these variables and their interactions are not strictly comparable to the other models in the Table. The difference in specification is minor and does not affect estimates of the other coefficients in the model.

The constant represents the baseline odds (not odds ratio) when categorical variables are at their reference values and age and time at values to which these were centered (age 50, year 2000).

IC indicates interview count.

(1) Sample: Column (A) is the main results model from Hale et al.[2] Column (B) shows the same model estimated on the Lee et al.[1] sample. The coefficients are similar, so sample differences are negligible. For comparability, the remaining regressions use the sample by Lee et al.[1]

(2) Measurement: Our measure of practice effects assumes that practice happens solely for the cognitive test taker; thus, the variable counts only self-interviews in which the respondent took a cognitive test. Lee et al.[1] count the number of participation waves regardless of whether the cognitive test score came from a self-interview or a proxy interview. Such a variable does not meaningfully reflect practice. Changing the measurement of practice from number of tests taken to number of participation waves attenuates the time trend odds ratio (OR) from 1.32 (95% confidence interval (CI) = 1.20, 1.45) (Column [B]) to 1.22 (95% CI = 1.11, 1.35) (Column [C]).

(3) Interaction: Lee et al.[1] argue the interaction accounts for differential experience for proxies versus self-interviews. Since interview count is highly correlated with time, the interaction is likely to pick up the differential proxy trend in cognitive function (as measured by proxy- vs. self-interviews), which stems from a change in how the Health and Retirement Study (HRS) administered interviews starting in 2006. It therefore aggravates multicollinearity in the model. The estimated time trend drops to 1.01 (95% CI = 0.87, 1.17) when this interaction is included (Column [D]). We agree there is a difference between proxies and self-interviews, but we consider our approach of excluding proxies from the practice accumulation variable and including separate proxy dummies preferable, as this provides meaningful practice definition with lower collinearity.

(4) Attrition: The measure of total number of interviews by Lee et al.[1] does not account for refreshment samples, thereby conflating recently recruited respondents with earlier recruits that attrited, and it does not account for the fact that only early recruits can accumulate a high number of interviews. This variable, therefore, does not capture propensity-to-attrit. It is when Lee et al.[1] introduce this variable that the time trend estimate becomes negative (Column [E]). This is the preferred model of Lee et al.[1] In addition to the flawed attrition variable, we view as problematic the measurement of practice effects (difference [2]) and the inclusion of the highly collinear interaction (difference [3]). We also find it concerning that the interview count coefficients change sign and suggest strongly increasing odds of dementia with increasing interview count. This pattern is counterintuitive and may signal collinearity problems (see difference [3]).

Column (F) shows what happens when the key propensity-to-attrit variable by Lee et al.,[1] total number of interviews (difference [4]), is added to a model that has a justifiable practice effects measure and no interaction between current interview count and proxy. This model estimates a time trend very close to our original result (OR = 1.30; 95% CI = 1.19, 1.43 vs. 1.32; 95% CI = 1.20, 1.45). In other words, the declining

trend result by Lee et al.[1] is not driven by the variable that is supposed to capture propensity-to-attrit, in contrast to what Lee et al.[1] imply, but by the other model choices that we critique above.

We conclude the empirical section by highlighting two issues related to both articles. Both are based on measures of cognitive function that include imputed values.[3] Although imputed values should not accrue practice effects, most imputations only concern a single component of the 27-point score, so removing imputed item responses is unlikely to impact results. Second, to capture all the waves in which the HRS consistently administered the four test components, we begin analysis at 1995/1996. In earlier waves, some groups of respondents were exposed to parts of the test. Both Lee et al.[1] and Hale et al.[2] ignore the potential practice effects emerging from these partial tests. It is unclear how accounting for partial tests affects the quality of the practice effect control, as it captures potential practice, but conflates different types of tests. Regardless, even if we re-count our practice effect variable under reasonable assumptions, the estimated OR for the time trend changes from 1.32 (95% CI = 1.20, 1.45) (Column [B]) to 1.25 (95% CI = 1.15, 1.36) (specifications in replication script[4]).

Despite our critique of the empirical implementation of the Lee et al.[1] model, their conceptual idea of jointly considering practice effects and selective attrition is important to consider. While we addressed this in two robustness checks—estimating joint models to test for selective mortality (Hale et al.[2] eAppendix; http://links.lww.com/EDE/B677) and stratifying estimation by interview count to purge the sample of other forms of sample selection (selective attrition and selective nonresponse)—Lee et al.[1] take a different approach. They begin by motivating their regression models using directed acyclic graphs (DAGs). However, it is unclear which of DAGs by Lee et al.[1] motivate their model choice. If it is the DAG in Lee et al.[1] Figure 2, it would be a mistake to regress measured cognition, Y*, on propensity-to-attrit, P, since P is caused by Y*. This may explain the counterintuitive coefficients for interview count, E, which are present even though, according to the DAG, this channel should be blocked. If their preferred DAG is a mixture of the left-hand side of Lee et al.[1] Figure 1 and of Lee et al.[1] Figure 2, the resulting graph is no longer acyclic (E=>Y*=>P=>E), so that the usual d-separation analysis no longer applies.

Our suggestion would be, if one tries to proxy for a time-constant propensity-to-attrit P, to conceive of it as a common cause of survey experience E and cognition Y (E<=P=>Y instead of E<=P<=Y*<=Y) while maintaining the E=>Y* edge, which would allow for the regression approach used in both articles. Still, proper adjustment for P would be needed for unbiased estimation of the effect of E. If P is correlated with time T, this is also true for unbiased estimation of the trend coefficient. Inclusion of a time-constant

propensity-to-attrit variable does not conclusively control for sample selection bias and may introduce additional biases—a complicated subject that goes beyond the scope of this short rejoinder.

We thank Lee et al.[1] for engaging with our article and for their commitment to open science. Replication code for all our analyses is available.[4] We hope that this rejoinder will facilitate further discussion on this important matter, as both teams agree that these multiple analytical challenges must be tackled to accurately assess dementia trends from the HRS data.

REFERENCES

1. Lee M, Halpern-Manners A, Warren JR. The prevalence of cognitive impairment is not increasing in the United States: a critique of Hale *et al*. (2020). *Epidemiology*. 2021;32:712–716.
2. Hale JM, Schneider DC, Gampe J, Mehta NK, Myrskylä M. Trends in the risk of cognitive impairment in the United States, 1996-2014. *Epidemiology*. 2020;31:745–754.
3. Fisher GG, Hassan H, Faul JD, Rodgers WL, Weir DR. *Health and Retirement Study Imputation of Cognitive Functioning Measures: 1992–2014*. University of Michigan Survey Research Center; 2017.
4. Hale JM, Schneider DC, Gampe J, Mehta NK, Myrskylä M. Stata replication code for Rejoinder to "The prevalence of cognitive impairment is not increasing in the United States: a critique of Hale *et al*. (2020)." *Open Scie Framework*. 2021. Available at: https://osf.io/h3d9s/.