Final Project

Keith Barnatchez

April 17, 2020

Abstract

I analyze the differences in baseline cognitive ability, as well as the rate of change in cognitive ability, in the offspring of centenarians vs. non-centenarians. Using data from a Boston University Telephone Interview for Cognitive Status (TICS) study, I find that centenarian offspring tend to exhibit higher levels of baseline cognitive function, after controlling for confounding factors such as age, family effects and smoking history. I find *no* evidence, however, that centenarian offspring enjoy slower rates of cognitive decline as they age. As an additional exercise, I use a mixture model to identify non-specified groups of study participants with signifcantly different changes in TICS scores over time, and am unable to find other subgroups of the data with significantly different rates of change in TICS over time. Given the finding of higher baseline TICS scores in centenarian offspring, these results suggest that further studies that link genetic and lifestyle factors to centenarians may be warranted.

1 Introduction

In efforts to limit the effects of chronic disease, centenarians—individuals who live to be at least 100 years—provide an opportunity to explore genetic and lifestyle factors that can increase longevity. One key metric associated with healthy aging is cognitive function, commonly measured by TICS (Telephone Interview for Cognitive Status) scores. To inform further studies in this growing field, it is important to understand whether centenarian offspring exhibit higher baseline intelligence levels, as well as slower rates of cognitive decline over time.

In this report, I analyze TICS data from a Boston University study to answer four main questions: 1) What characteristics differ between descendants of centenarians vs. non-centenarians? 2) Does cognitive function at baseline differ between centenarian vs. non-centenarian offspring after adjusting for differential factors? 3) Does that rate of change over lifetime in cognitive function differ between centenarian vs. non-centenarian offspring? and 4) Do there exist other, non-hypothesized groups of individuals that differ in their rate of change of cognitive function? Answers to these questions will inform whether additional—and potentially costly—research on the genetic and lifestyle factors of centenarians is warranted in efforts to reduce cognitive decline.

2 Methods

In this section, I describe the data collection process and modeling methods used to address my main research questions.

2.1 Data

The data were collected from two studies of longevity and healthy aging conducted at Boston University. In these studies, families were enrolled and participating members were given a TICS test: a series of questions conducted by phone to assess cognitive ability. Respondents took as many five TICS tests, typically dispersed by at least two years, though many respondents did not take all five tests due to death and other factors.[1] The unit of observation in the data is a person, who has a corresponding ID and family ID. Along with their TICS score, each person has measurements on factors such as age, gender, history of smoking, and history of cancer. For the full set of variables in the data and their summary statistics, see the appendix.

To examine which characteristics differ significantly between offspring of centenarians vs. non-centenarians, I sequentially test for significant differences between the two groups for each variable. For an arbitrary variable X_i where $i \in \{\text{N}, \text{C}\}$ denotes either centenarian (C) or non-centenarian (N) offspring, I specify a simple Bayesian regression model when the X_i is continuous:

$$X_i \sim N(\mu_i, \tau) \qquad \mu_i = \mu_0 + \theta \times \text{Centenarian}_i$$

with priors

$$\mu_0 \sim N(0, 10^6) \qquad \tau \sim \text{Gamma}(1, 1) \qquad \theta \ \sim N(0, 10^6)$$

and a simple logistic model when $X_i \in \{0, 1\}$:

$$X_i \sim \text{Binom}(p_i, 1) \qquad \text{logit}p_i = p_0 + \theta \times \text{Centenarian}_i$$

with priors

$$p_0 \sim N(0, 10^6) \qquad \theta \sim N(0.10^6)$$

I use 95% credible intervals to determine which variables have significant differences between centenarian offspring vs. controls. Using this approach, I find that the following variables are significantly different: sex, age of last contact, BMI, smoking status, aspirin use, stroke, coronary artery disease, diabetes and hypertension history, and status at last contact (alive or deceased).

[1] I discuss the challenge this poses for modeling in the following section.

Figure 1 displays the distributions of the first through fifth TICS test scores, split by the centenarian and control groups. While it appears that the non-centenarian scores become relatively more variable over time (which could suggest a greater rate of decline in TICS scores over time), it is difficult to discern whether thse differences exist from the raw data. Similarly, in the left panel of Figure **??** it is difficult to determine whether there are differential rates of change in TICS between centenarian and non-centenarian offspring. In the following section, I present a model-based approach to help identify potential differences.

2.2 Main Model

To examine whether there are significant differences in baseline TICS and/or the rate of change in TICS over time, I specify the following Bayesian hierarchical model:

$$\begin{aligned}
&\text{TICS}_{ijt} \sim N(\mu_{ijt}, \sigma) \\
&\mu_{ijt} = \alpha_j + \beta_1 \text{C}_i + \beta_2 \text{BaselineAge}_{it} + \beta_3 \Delta\text{Age}_{it} + \beta_3(\text{C}_i \times \Delta\text{Age}_{it}) + \sum_{k=1}^{K} \gamma_k X_{kijt}
\end{aligned}$$

with uninformative priors (where letter subscripts imply the same prior for all indices):

$$\beta_i \sim N(0, 10^6),\ \alpha_j \sim N(m_j, \tau),\ \gamma_k \sim N(0, 10^6),\ m_j \sim N(0, 10^6),\ \tau \sim \text{Gamma(1,1)},\ \sigma \sim \text{Gamma}(1, 1)$$

where TICS_{ijt} is the TICS score of person i from family j observed at time t. X_{kijt} is the k-th additional control variable (for the corresponding person, family and time), C_i denotes whether individual i has is a centenarian offspring, and ΔAge_{it} denotes how many years individual i has aged in time t relative to their baseline age.[2] The coefficients of interest are β_1 (the baseline effect) and β_3 (the rate effect). Importantly, the inclusion of family-level random intercepts α_j helps to account for time-invariant, family-specific factors that may otherwise be attributed to centenarian effects.

For my main set of additional controls, I use the variables that are significantly different between centarian and non-centarian offspring at baseline, as listed earlier. This list of controls includes BMI, which is missing for some individuals. To account for this information loss, I impute the missing BMI values by modeling its conditional mean as a linear function of the other controls in the main model:

$$\text{BMI}_i \sim N(\theta_i, \nu) \quad \theta_i = a_0 + \sum_{k=1}^{K} a_k x_k$$

with priors (where letter subscripts imply the same prior for all indexes):

$$a_i \sim N(0, 10^6) \quad \nu \sim \text{Gamma}(1, 1)$$

It should be noted that not every individual completed all 5 follow-up surveys. Causes for response attrition include death and inability to follow up with contacts. While we could attempt to model the missingness in TICS explicitly, it is more suitable to think of missing TICS data as a censoring issue, and censored modeling methods require more nuance and care to present than is appropriate for this brief report. Thus, I do not account for missing follow-up TICS scores in my analysis, and remove them from the data.

2.3 Mixture Model

The model presented above allows one to analyze whether there are significant level and/or rate effects from having centenarian parents, but it makes it difficult to identify if there are other groups within the data with significantly different rates of change of TICS that were not considered a priori. Rather than arbitrarily test each group variable for significant differences in TICS rates of change, I look for such groupings with a model-based approach. I implement a simple, 2-state mixture model that replaces the centenarian group with a latent variable ε, which can equal 0 or 1:

[2] Using this variable, instead of unadjusted age, allows us to interpret a respondent's first TICS score as their baseline score rather than the score they received at a particular age.

$$\text{TICS}_{ijt} \sim N(\mu_{ijt}, \sigma_j)$$

$$\mu_{ijt} = \beta_0 + \alpha_j + \beta_1 \varepsilon_i + \beta_2 \text{BaselineAge}_i + \beta_3 \Delta\text{Age}_{it} + \beta_3(\varepsilon_i \times \Delta\text{Age}_{it}) + \sum_{k=1}^{K} \gamma_k X_{kijt}$$

The priors are the same as before (see the main model above), with an additional prior on the probability parameter θ: $\varepsilon_i \sim \text{Binom}(\theta, 1),\ \theta \sim \text{Gamma}(1, 1)$

I fit both the main and mixture models with Gibbs sampling. I take care to make sure proper convergence of all constructed Markov Chains – more details on convergence diagnostics are included in the RMarkdown HTML file.

3 Results

I first turn my attention to the main model's results in Table 2. I report the median and 95% credible intervals for the main model parameters (not including the coefficients used to impute missing BMI values). The estimated baseline effect of having a centenarian parent, β_{C}, is 0.53 additional TICS points relative to the control group, all else in the model equal.[3] The 95% credible interval ranges from 0.16 to 0.91, suggesting that this positive effect is statistically significant.

As the 95% credible interval for $\beta_{\text{C}\times\text{Age}}$ contains 0, we do not have sufficient evidence that there are significantly different rates of change in TICS between centenarian and non-centenarian offspring. Unsurpisingly, I find that TICS scores tend to decrease with age, with the posterior median of β_{Age} being -0.189, suggesting a expected decline in TICS of 0.189 points for each year an individual ages after taking the baseline test – this effect is also statistically significant.

The mixture model appears to arbitrarily split the trajectories of TICS scores in half, as the probability parameter θ of being in group 1 has a posterior median of 0.48 (see Table 3) – this is likely because the raw trajectories are highly mixed. The right panel of figure 2 shows TICS trajectories, colored by group assignment, and we see the model effectively splits the trajectories in half when grouping, at roughly the mean TICS score. Even with the group assignments being questionable, the posterior credible interval for $\beta_{\varepsilon\times\text{Age}}$ contains 0, suggesting the rates of change in TICS between the two groups are not significantly different.

4 Discussion

I find that centenarian offspring tend to exhibit higher levels of baseline TICS scores – the posterior median increase of 0.53 TICS points represents a 3.8 percent increase above the baseline average TICS score. Importantly, this effects persists after controlling for family-level factors with the inclusion of random effects. While I find no evidence of different rates of change in TICS scores, this positive finding suggests that further studies on centenarians that examine associated genetic and lifestyle factors could be warranted given their potential relationship to increased cognitive function. I was not able to identify other groups within the data with significantly different rates of change in TICS, nor did I find that centenarian offspring had significantly different rates of change in TICS scores. Further studies with richer data are required to identify potential factors that influence cognitive decline. Though I do not find evidence for it here, researchers should not rule out the possibility that centenarians, or other unidentified groups, exhibit relatively lower rates of cognitive decline.

Of course, there are limitations to my analysis – I am unable to account for censoring of TICS responses, and it is possible that there are confounding variables that were omitted from the model due to lack of data availability. As a result, these results should not be interpreted as causal estimates and rather as measures of association. Further, more nuanced analysis is required to learn more about the relationship between centenarians and cognitive function, though researchers should view the results from this short analysis as encouraging.

[3] To aid the reader, I refer to coefficients from this point on with subscripts for their corresponding variable. E.g. I refer to β_1 as β_{C} and β_2 as $\beta_{\text{Baseline Age}}$.

5 Tables and Figures

Table 1: Summary statistics of baseline variables

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Sex	762	0.579	0.494	0	0	1	1
Centenarian offspring	762	0.646	0.479	0	0	1	1
Age at enrollment	762	70.678	7.396	39	66	76	90
BMI	724	27.243	4.828	17.782	24.013	29.835	51.181
Smoking history (1 if yes, 0 if no)	762	1.570	0.495	1	1	2	2
Aspirin use history	761	1.466	0.499	1.000	1.000	2.000	2.000
Stroke history	762	1.026	0.160	1	1	1	2
Diabetes history	762	1.076	0.265	1	1	1	2
Hypertension history	762	1.352	0.478	1	1	2	2
Coronary artery disease history	762	1.059	0.236	1	1	1	2
Cancer history	762	1.252	0.434	1	1	2	2
Years of education	762	15.792	2.903	8	14	18	26
Baseline TICS score	760	13.908	4.251	2.000	11.000	16.000	26.000

Figure 1: TICS distributions over survey rounds

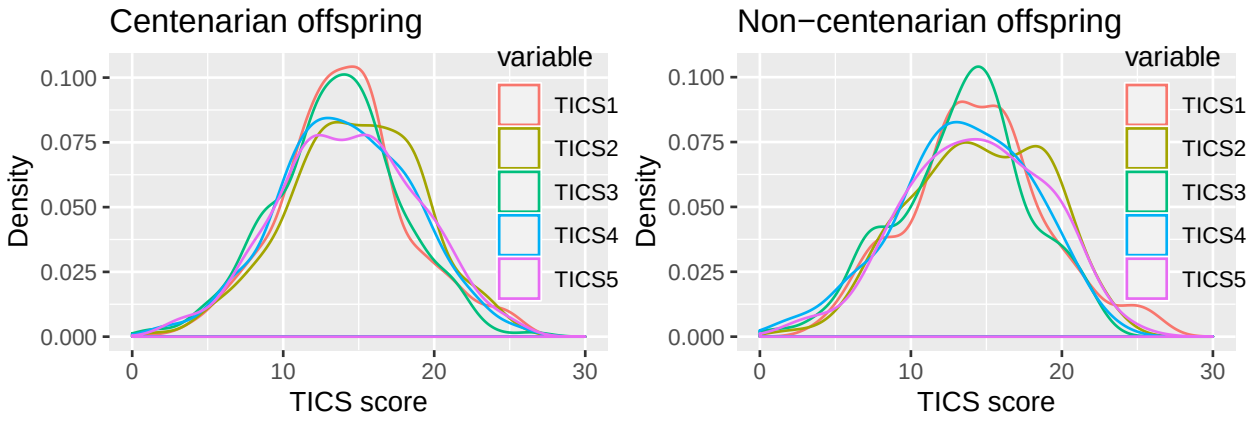

Figure 2: TICS score trajectories: centenarian offspring vs. controls

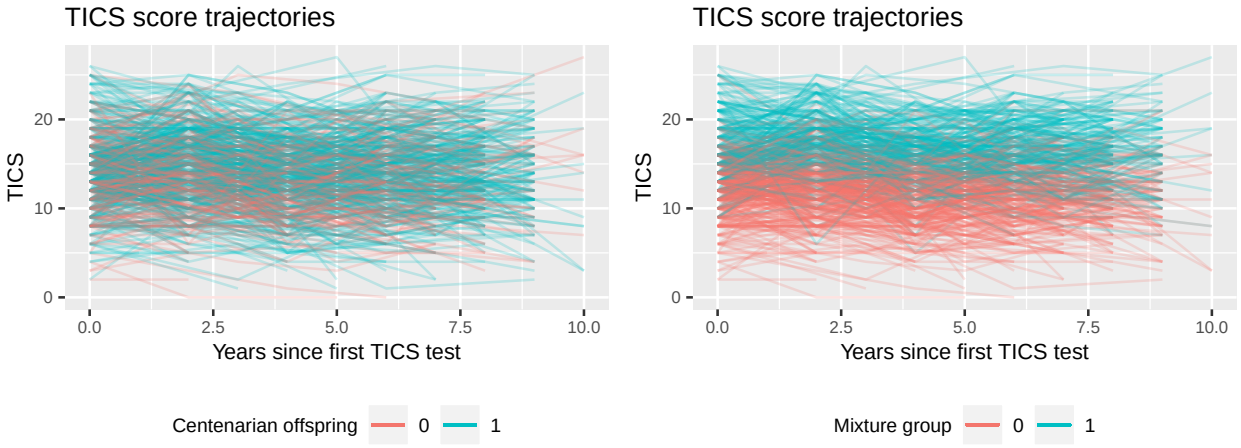

Table 2: Main model results

	Posterior Median	95 percent CI
$\beta_{\Delta\text{Age}}$	-0.1887	(-0.262,-0.112)
$\beta_{\text{Baseline Age}}$	-0.1939	(-0.224,-0.163)
$\beta_{\text{Centenarian}}$	0.5276	(0.161,0.909)
$\beta_{\text{Centarian}\times\Delta\text{Age}}$	0.0657	(-0.0254,0.153)
γ_{Male}	1.7549	(1.44,2.07)
γ_{BMI}	-0.0878	(-0.127,-0.049)
γ_{Smoke}	0.3187	(-0.0569,0.72)
γ_{Aspirin}	0.2452	(-0.0693,0.578)
γ_{Stroke}	-1.2575	(-2.47,-0.259)
$\gamma_{\text{CoronaryDisease}}$	-0.1459	(-0.874,0.612)
γ_{Diabetes}	0.0135	(-0.734,0.763)
γ_{Deceased}	-1.6872	(-2.2,-1.16)

Table 3: Mixture model results

	Posterior Median	95 percent CI
$\beta_{\Delta\text{Age}}$	-0.1181	(-0.178,-0.0592)
$\beta_{\text{Baseline Age}}$	-0.2177	(-0.248,-0.189)
β_{ε}	-4.5051	(-4.77,-4.23)
$\beta_{\varepsilon\times\Delta\text{Age}}$	-0.0520	(-0.141,0.0417)
γ_{Male}	1.7166	(1.26,2.19)
γ_{BMI}	-0.1024	(-0.14,-0.0663)
γ_{Smoke}	0.9438	(0.587,1.33)
γ_{Aspirin}	0.5429	(0.161,0.941)
γ_{Stroke}	-1.3515	(-2.28,-0.224)
$\gamma_{\text{Coronary Disease}}$	0.0377	(-0.834,0.924)
γ_{Diabetes}	-0.8563	(-1.71,-0.125)
γ_{Deceased}	-1.7542	(-2.34,-1.2)
θ	0.4823	(0.405,0.555)