Preregistration

# Personality Predictors of Dementia Onset and Progression: A Mega-Analysis

Emorie D. Beck[1], Daniel K. Mroczek[1,2], Eileen K. Graham[1]

[1] Feinberg School of Medicine

[2] Northwestern University

*18. August 2020*

## Study Information

| | |
|---|---|
| **Title** | Personality Predictors of Dementia Onset and Progression: A Mega-Analysis |
| **Description** | Personality traits are relatively stable, dispositional patterns that differentiate people from one another (Roberts, Wood, & Caspi, 2008). Personality traits are robust predictors of many life outcomes (Beck & Jackson, 2020), including health outcomes such as disease onset (Weston et al., 2015), terminal cognitive decline (Wilson et al., 2015) and mortality risk (Mroczek & Spiro, 2007, Graham et al., 2017). They also are critical predictors of important factors that lead to later disease or early mortality, most notably health behaviors (Turiano et al., 2014) and physician adherence (Hill & Roberts, 2011). In fact, this is one mechanism by which psychological dimensions such as well-being or personality influence more distal health outcomes: via mediating factors such as lifestyle factors. Similar processes may operate with respect to post-dx cognitive decline among those with dementia. |

Those with particular personality- or well being-based predisposing factors, may have slower rates of decline, meaning less impairment and better functioning for a longer length of time even after the onset of dementia.

---

**Research Question**

1. Do personality and subjective well-being indicators (positive affect, negative affect, and satisfaction with life) predict dementia onset?
2. Do personality and subjective well-being indicators (positive affect, negative affect, and satisfaction with life) predict disease progression among those diagnosed with dementia?

# Design Plan

---

**Study type**

**Observational Study**. Data is collected from study subjects that are not randomly assigned to a treatment. This includes surveys, natural experiments, and regression discontinuity designs.

---

**Study design**

### Health and Retirement Study (HRS)

The Health and Retirement Study (HRS; Juster & Suzman, 1995) is an ongoing longitudinal study of households in the United States. These data are available at https://hrs.isr.umich.edu by creating a free account.

Participants were recruited from more than 35,000 individuals from the financial households of individuals born between 1931 and 1941 in the US. Data have been collected biannually since 1992. The latest data release includes data up to 2016. On average, 10,000 individuals are sampled each wave More information on the HRS can be found at https://hrs.isr.umich.edu/documentation/survey-design, but, in short, the HRS is a nationally representative sample of adults over 50 in the US. It is critical to note that the HRS samples households of the original cohort and follows individuals and their spouses or partners until their death.

Sample size varies by year, ranging from approximately 7,500 (2014) to 15,500 (1992). (https://hrs.isr.umich.edu/sites/default/files/biblio/ResponseRates_2017.pdf). This provides 99% power to detect a zero-order correlation effect size of ~.04, two-tailed at alpha .05.

### RUSH Memory and and Aging Project (RUSH-MAP)

The RUSH Memory and Aging Project (RUSH-MAP) is an ongoing longitudinal study that began in 1997. These data are available, through application from https://www.radc.rush.edu/requests.htm.

Participants who were 65 and older were recruited from retirement communities and subsidized senior housing facilities throughout Chicagoland and northeastern Illinois beginning in 1997. Data are collected annually, and all participants are organ donors. Additional participants are recuited each year. Additional information and documentation on the data can be found at https://www.radc.rush.edu/docs/var/variables.htm.

Sample sizes vary by year, ranging from 52 (1997) to 2205 participants including 884 deceased participants with autopsy data (2019, 2020). This provides 99% power to detect a zero-order correlation effect size of ~.10, two-tailed at alpha .05.

### RUSH Religious Orders Study (ROS)

The RUSH Religious Orders Study (ROS) is an ongoing longitudinal study that began in 1994. These data are available, through application from https://www.radc.rush.edu/requests.htm.

Older (65 and above) Catholic nuns, priests, and brothers with no prior dementia diagnosis and who agreed to annual evaluations and eventual organ donation were recruited from more than 40 groups across the United States. Additional participants are recuited each year. Additional information and documentation on the data can be found at https://www.radc.rush.edu/docs/var/variables.htm.

Sample sizes vary bt year from 353 participants (1994) to 1487 participants, including 797 deceased participants with autopsy data (2019, 2010). This provides 99% power to detect a zero-order correlation effect size of ~.11, two-tailed at alpha .05.

### Swedish Adoption Twin Study of Aging (SATSA)

The Swedish Adoption Twin Study of Aging (SATSA) is a longitudinal study of twin pairs from the Swedish Twin Registry that began in 1984. Data are available through the ICPSR database at https://www.icpsr.umich.edu/web/ICPSR/studies/3843.

All twin-pairs on teh Swedish Twin Registry who were separated at an early age were invited to be a part of the study in 1984. A control sample of twins reared together were also included. Additional waves of all participants were collected in 1987, 1990, 1993, 2004, 2007, 2010, 2012, and 2014. More information, including codebooks, scales, and variable search functions can be found at https://www.maelstrom-research.org/mica/individual-study/satsa/#.

Sample sizes vary by wave, ranging from 2018 participants at baseline (1984) to 379 participants (IPT7). Given that the target measures were collected at baseline, this provides 99% power to detect a zero-order correlation effect size of ~.10, two-tailed at alpha .05.

### ADRC Memory and Aging Project (ADRC-MAP)

The Alzheimer Disease Research Center Memory and Aging Project (ADRC-MAP) is an ongoing longitudinal study of memory and Alzheimer's Disease that began in 1979. Data are available on a study-by-study basis through application from https://knightadrc.wustl.edu/Research/ResourceRequest.htm.

Participants were recruited from the Charles and Joanne F. Knight Alzheimer's Disease Research Center at Washington University in St. Louis as part of an ongoing study of disease progression. The current study uses a subset of approximately 1200 of these participants who completed personality surveys as part of a sub-study (see Duchek et al., 2019). More information on the study can be found at https://knightadrc.wustl.edu/Research/PDFs/Clinical%20Core%20list%20of%20measures.pdf.

Sample sizes vary over time, from approximately 400 to 1200. This provides 99% power to detect a zero-order correlation effect size of ~.15, two-tailed at alpha .05.

**Baltimore Longitudinal Study of Aging (BLSA)**

## Sampling Plan

**Existing data**

**Registration prior to accessing the data**. As of the date of submission, the data exist, but have not been accessed by you or your collaborators. Commonly, this includes data that has been collected by another researcher or institution.

**Registration prior to analysis of the data**. As of the date of submission, the data exist and you have accessed it, though no analysis has been conducted related to the research plan (including calculation of summary statistics). A common situation for this scenario when a large dataset exists that is used for many different studies over time, or when a data set is randomly split into a sample for exploratory analyses, and the other section of data is reserved for later confirmatory data analysis.

**Explanation of existing data**

The first author, who will be conducting all analyses, has previously worked with data from the Health and Retirement Study, including key personality data. However, she has not previously worked with dementia data from HRS. She has not previously worked with or accessed RUSH MAP and ROS, ADRC Map, SATSA, and BLSA data. All core variables, data cleaning procedures, and analytic procedures for each of these studies are included with the present preregistration.

**Sample size rationale**

Sample sizes vary across study and waves (see above). But each of these studies have at least 1000 participants, which provides 99% power to detect correlations of at least .14, two-tailed alpha at .05. And overall estimates will have considerably higher power.

**Stopping rule**

The authors did not collect these data and most of these studies are still enrolling participants. Sample sizes will be determined by data sent to the authors shortly after preregistration.

## Variables

**Measured variables**

There will four categories of data:

(1) Personailty will be measured using Big Five Personality scales. Well-being measures will be separate measures of positive affect, negative affect, and satisfaction with life separately, as available in each study.

(2) Cognitive and physical symptom dementia diagnoses will be measured using self-report diagnoses, CDR's, and clinician diagnoses.

Post-mortem dementia diagnoses and severity will be measured using a subset of vascular, neuropathological, cerebrospinal, and imaging indices, may of which are based on the National Alzheimer's Coordinating Center's Neuropathology Data Form:

- Braak stage (0 to 6)
- CERAD (1 = definite to 4 = no AD)
- Overall Amyloid Beta Score (average cross-region percent area of cortex oocupied by amyloid beta)
- Diffuse plaque burden (average of scaled regional counts)
- Neuritic plaque burden (average of scaled regional counts)
- Neuronal neurofibrillary tangle density (mean density across regions per $mm^2$)
- Neuronal neurofibrillary tangle burden (average of scaled regional counts)
- Lewy Body Disease (4 level; 0 = none, 1 = nigral-predominant; 2 = limbic-type; 3 = neocortical-type) - Gross Cerebral Infarcts (0 = No, 1 = Yes)
- Gross Cerebral Microinfarcts (0 = No, 1 = Yes)
- Cerebral Atherosclerosis (0 = none to 3 = severe)
- Cerebral Amyloid Angiopathy (0 = none to 3 = severe)
- Arteriolosclerosis (0 = none to 3 = severe)
- Hippocampal Sclerosis (0 = No, 1 = Yes)

(3) Cognitive Ability (and decline) will be measured using sets of tests specific to each study, many of which are drawn from the Wechsler Adult Intelligence Scale.

(4) Background variables (age, gender, SES, race, education)

More detail on the scale, recoding, and summarizing of these data are available in the codebook attached with this preregistration.

---

**Indices** Personality and well-being measures will be composited into trait scores preserving the original scale in each data set. Then, personality scores in each data set will be converted to Percentages Of the Maximum Possible score (POMP) in the mega-analytic procedure (Cohen, Cohen, Aiken, & West, 1999). Unlike standardization procedures, that have a mean of zero and unit variance and can be misleading when data are skewed, POMP does not rescale sample variance based on the observed data, which overly relies on deviations from the mean. Instead, POMP relies on the ratio between the difference between a score and the minimum and the maximum and minimum, or

$\text{POMP} = \frac{observed - minimum}{maximum - minimum}$.

More information on the exact scales used to create the variables are in the table below.

## Analysis Plan

The analysis plan will be broken into parts to answer different questions associated with this study.

Following pre-registration of this study, all data will be downloaded from study websites or received directly from data maintainers for each study. All target variables, as well as their rescaling and cleaning procedures, have been preregistered with this study.

After all data are received, data for each study will be cleaned separately by the first author. These cleaning procedures can be roughly broken into pulling, rescaling, and/or compositing (1) background variables at baseline, (2) baseline personality and well-being (henceforth, baseline personality), (3) longitudinal cognitive ability scores (to measure cognitive decline), (4) self-reported or clinician dementia diagnoses, (5) biomarkers of post-mortem dementia (tau protein levels and $A\beta_{42}$ and $A\beta_{40}$).

Once each data set is prepared, data sets across studies will be combined and scales will be harmonized as needed (e.g., POMP for personality data will be calculated

separately for each test due to differences in data missingness across models).

Once data are combined, we will address the two main research questions.

1. Do personality and subjective well-being indicators (positive affect, negative affect, and satisfaction with life) predict dementia onset?

First, to test whether personality predicts dementia onset, we will run two separate models because diagnosis is a categorical (0 = no diagnosis, 1 = diagnosis) while biomarkers are continuous.

For diagnoses, we will run a series of bayesian multilevel logistic regression models, while for biomarkers we will run a series of bayesian multiple regression models. For both the basic model equation is the same:

Level 1: $Y_{ij} = \beta_{0j} + \beta_{1j} * P_{ij} + \varepsilon_{ij}$ Level 2:
$\beta_{0j} = \gamma_{00} + u_{0j}$
$\beta_{1j} = \gamma_{10} + u_{1j}$,

For the logistic regressions, $\gamma_{00}$ is the average log odds of experiencing the outcome across all studies and $\gamma_{10}$ multiple of log odds change associated with a one-unit change in the percentage of the maximum of the possible (POMP) personality score. All results will be presented both as log odds and as odds ratios (OR) with 95% credibility intervals (CI). $u_{0j}$ indicates the difference between the average estimate of log odds of experiencing an outcome and the estimate for each study (i.e. the study-specific estimate of the log odds of each outcome), and $u_{1j}$ indicates the difference between the average multiple of log odds associated with a one unit change in POMP personality score and the estimate for each study (i.e. the study-specific estimate of the personality-outcome relationship). Each of these will be presented as forest plots showing both study-specific and average effects.

For the regressions predicting biomarkers from personality, $\gamma_{00}$ is the average level of each biomarker across all studies and $\gamma_{10}$ change in each biomarker associated with a one-unit change in the percentage of the maximum of the possible (POMP) personality score. All results will be presented as unstandardized regression coefficients with 95% credibility intervals (CI). $u_{0j}$ indicates the difference between the average estimate of biomarker levels and the estimate for each study (i.e. the study-specific estimate average biomarker levels), and $u_{1j}$ indicates the difference

between the average biomarker levels associated with a one unit change in POMP personality score and the estimate for each study (i.e. the study-specific estimate of the personality-biomarker relationship). Each of these will be presented as forest plots showing both study-specific and average effects.

2. Do personality and subjective well-being indicators (positive affect, negative affect, and satisfaction with life) predict disease progression among those diagnosed with dementia?

Next, we will examine whether personality predicts decline following diagnoses. To do this, we will create two separate sub-samples for each study, one based on clinical diagnoses and one based on autopsy diagnoses. Next, we will estimate slopes of cognitive variables and test whether personality moderates the these trajectories.

To do so, we will estimate a series of Bayesian cross-classified random slope piecewise multilevel models. The cross-classified random slope model is an extension of the basic piecewise growth model typically used in examination of socialization effects below:

Level 1: $Y_{ij} = \beta_{0j} + \beta_{1j} * prewave_{ij} + \beta_{2j} * postwave_{ij} + \epsilon_{ij}$
Level 2:
- $\beta_{0j} = \gamma_{00} + \gamma_{01} * personality_j + u_{0j}$
- $\beta_{1j} = \gamma_{10} + \gamma_{11} * personality_j + u_{1j}$
- $\beta_{1j} = \gamma_{20} + \gamma_{21} * personality_j + u_{2j}$

where $\gamma_{00}$ is the average cognitive ability score of those with the lowest score in each personality measure at baseline, $\gamma_{01}$ is the average difference in baseline cognitive ability associated with a one unit increase in personality, $\gamma_{10}$ is the average slope of cognitive ability change of those with the lowest score in each personality measure at baseline, and $\gamma_{11}$ is the change in average slope of cognitive ability associated with a one unit change in personality. Most critically, however, are the post-dementia coefficients, $\gamma_{20}$ and $\gamma_{21}$, which capture the difference between pre- and post-event levels and change. Specifically, $\gamma_{20}$ is the difference between pre- and post-dementia average cognitive ability change among those iwth the lowester score in each personailty measure at baseline, and $\gamma_{21}$ is the difference in pre-post trajectory change associated with a one unit change in personality. $u_{0j}$ captures the difference between the overall average baseline cognitive ability score and an

individual's score, $u_{1j}$ captures the differences between the overall average cognitive ability change and an individual's change, and $u_{2j}$ captures individuals' deviations from teh difference in pre-post trajectory change.

The cross-classified random slope model extends this by added a second, non-hierarchical random intercept and slope to the basic socialization growth model as below:

Level 1: $Y_{ij} = \beta_{0j} + \beta_{1j} * prewave_{ij} + \beta_{2j} * postwave_{ij} + \epsilon_{ij}$

Level 2:

- $\beta_{0j} = \gamma_{00} + \gamma_{01} * Personality_j + u_{0j,study} + u_{0j,sid}$
- $\beta_{1j} = \gamma_{10} + \gamma_{11} * Personality_j + u_{1j,study} + u_{1j,sid}$
- $\beta_{2j} = \gamma_{20} + \gamma_{21} * Personality_j + u_{2j,study} + u_{2j,sid}$

The notation here has been somewhat forced into a framework that matches the basic growth model for ease of interpretation. But essentially, it estimates cross-classified random effects for each study-person combination. These are independent in the sense that they are necessarily uncorrelated. Given these effects, we can estimate and report study-specific and person-specific intercepts, group differences in intercepts, slopes, and cross-level interactions between personality and cognitive ability change (i.e. does personality predict change).

All results will be presented both as unstandardized estimates (in terms of POMP) with 95% credibility intervals (CI). Each effect will be presented as forest plots showing both study-specific and average effects or spaghetti plots showing person-specific effects.

## Statistical models

**Transformations**

All data recoding, compositing, and transformations are meticulously documented in the codebook attached with this preregistration. This codebook will be directly read into R for the purposes of rescaling and recoding variables to ensure maximal transparency and reproducibility.

Personality will be scaled as POMP across studies.

| | |
|---|---|
| **Inference criteria** | Inference tests will all be based around the Bayesian 95% credibility interval (CI) of the effect. Intervals that do not overlap with 0 will be considered statistically significant. |
| **Data exclusion and missing data** | Every effort will be made to use the maximum amount of data possible. However, participants who are missing key personality, well-being, dementia diagnosis, biomarkers, cognitive ability scores, or demographic variables will be dropped in analyses involving those variables. Where possible, they will still be included in other tests.<br><br>Participants who have no record of any personality characteristic or dementia diagnosis (positive or negative) be dropped completely. |
| **Exploratory analyses (optional)** | Enter your response here. |

## Other

| | |
|---|---|
| **Other (Optional)** | Enter your response here. |

## References