Preregistration

Selection and Socialization: A Propensity Score Matched Study of Personality and Life Events

Emorie D. Beck[1], Jule Specht[2], Wiebke Bleidorn[3], Joshua J. Jackson[1]

[1] Washington University in St. Louis

[2] Humboldt-Universitaät zu Berlin

[3] University of California, Davis

25. January 2018

Introduction

Project Focus This project uses longitudinal data to examine the role of personality in predicting a set of life events, as well as personality changes following these experiences.

Hypotheses

Predictions and Rationales We have two categories of hypotheses:

1. *selection effects:* which of and to what extent do the Big 5 predict the experience of specific life events?

2. *socialization effects:* Controlling for selection effects, which of and to what extent do the Big 5 change following the experience of life events?

Previous work (Specht, Egloff, & Schmuckle, 2011) has examined the selection into and socialization effects of life events on the Big 5, but the present study extends their work in the following ways:

1. Inclusion of a third wave of personality data collected in 2013.
2. Inclusion of additional life event data collected annually from 2010 to 2013.
3. The use of multiple imputation to deal with missing data.
4. The use of propensity score matching to equate individuals who did or did not experience each life event on a number of background characteristics (see Table 1).
5. The use of bayesian structural equation modeling.

In other words, rather than directly examining selection effects, we using propensity score matching to eliminate them so we can examine socialization effects of life events independently of baseline differences in various characteristics, including the personality characteristics for which we are examining change.

To date, propensity score matching and weighting have been applied in a handful of studies in personality on a restricted set of life events, including entering into the military (Jackson, Thoemmes, Jonkmann, Lüdtke, & Trautwein 2012), transitioning to parenthood (van Scheppingen et al., 2016), entering into romantic relationships (Wagner, Becker, Lüdtke, & Trautwein, 2015), and moving into managerial positions (Niess & Zacher, 2015), so what we would expect based on the studies utilizing propensity score matching is extremely limited. Thus, we largely base our predictions on previous findings concerning the relationship between personality change and life events using the GSOEP sample.

Socialization Effects

Two main theories guide predictions about how personality changes, regardless of whether life events occur: (1) an essentialist perspective that argues that personality is genetically based and biologically determined, and a contextualist or social cognitive perspective that argues about the importance of social and environmental influences. The contextualist, but not the essentialist, perspective predicts personality change.

Given evidence that personality does change (or differ) throughout the lifespan, both longitudinally (Roberts et al., 2006) and cross-sectionally (Soto et al., 2011; Specht et al., 2011; Beck, Condon, & Jackson, in prep), and that longitudinal trait changes correlate with longitudinal state changes in personality (Beck & Jackson, in prep) we proceed from an enivornmental perspective – that is, we expect that personality will change. Specifically, according to the model of person-environment transactions (Roberts et al., 2008), the *Plasticity Principle* suggests that personality traits are changeable in any environment at any age. Change can occur and and that can be at any time throughout the life course. These general life course changes are termed normative maturation and are outlined below.

- **Extraversion**: the social dominance facet tends to curvilinearly increase until around age 50 (Roberts, Walton, & Viechtbauer, 2006) after which it tends to decline (Specht et al., 2011), while the social vitality facet tendes to slightly increase in young adulthood and then slightly decline in older adulthood across the lifespan. Because the items used in the present study reflect social vitality (sociability and energy) more than social dominance (self confidence, independence, and dominance), we expect to see overall increases in Extraversion in early adulthood followed by slight declines later in life.
- **Agreeableness** increases across the lifespan (Roberts et al., 2006; Specht et al., 2011).
- **Conscientiousness** increases throughout the lifespan (Roberts et al., 2006). In this sample, age differences suggest C increases in young adulthood, is relatively stable in middle adulthood, and decreases slightly in middle to older adulthood (Specht et al., 2011).
- **Neuroticism (Emotional Stability)** decreases throughout young and middle adulthood, particularly during young adulthoood, and increases slightly in

older adulthood (Roberts et al., 2006; Specht et al., 2011).

- **Openness** longitudinally shows a upside down u shaped curve, increasing in young adulthood and decreasing in older adulthood (Roberts et al., 2006). In this sample, openness cross-sectionally decreased throughout the lifespan, particularly in young and older adulthood (Specht et al., 2011).

Group differences in slopes (Life Event Socialization Effects)

The mere observation of normative change does not explain why it occurs. To explain change, we again turn to three principles of the model of person-environment transactions (Roberts et al., 2008) to provide possible mechanisms of change: the role continuity principle, the social investment principle, and the corresponsive principle (see below).

- The *Role Continuity Principle* suggests that roles, rather than events, change personality. In other words, we can expect life events that are associated with role changes to have a greater influence on personality (socialization effects). For example, beginning starting a first job is likely to come with increased responsibilities and different constraints on an individuals' time in order to fulfill the role of an employee, while an internal promotion within a job may result in minimal role change and is less likely to impact personality.
- The *Social Investment Principle* suggests that investment in social institutions is the primary vehicle through which people take on and invest in new roles. It is only through an *investment* in a role that change will occur. Notably, social institutions include relationships, career roles, or any other entity with a set of norms.
- The *Corresponsive Principle* suggests that people select into life events and roles that align with their personality characteristics – that is, the relationship between personality and life events is cyclical. In other words, when slopes are correlated with intercepts, the relationship should be positive.

Together, this suggests that socialization and selection effects are not independent – that is, baseline differences may masquerade as socialization or selection effects of a personality trait, when in fact, these effects are explained by other variables. We propose five possibilities for how life events may affect the slopes of individuals (see also Figure 3).

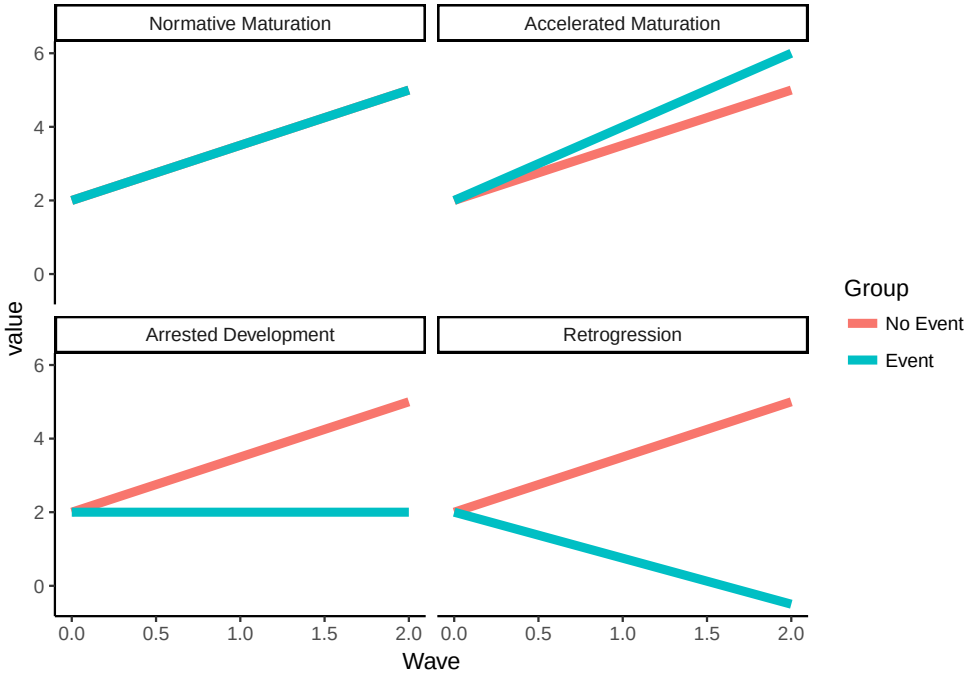
Hypotheses and Rationale:

2a. **Normative maturation.** Once accounting for baseline characteristics, people will experience similar development regardless of whether they experienced a life event (i.e., no group differences in slopes). Evidence for this from the Big 5 comes from longitudinal studies of the transition to parenthood (van Scheppingen et al., 2016) and military training (Jackson et al., 2012), who found minimal change in the Big 5 among people who did or did not experience these events.

2b. **Accelerated maturation.** People who experience a life event will show steeper positive changes in personality than those who do not, regardless of whether they had baseline advantages over those who did not. This would provide evidence for the importance of social roles – those who take those roles on sooner show change that resembles normative change but earlier in the life course. Evidence for this comes from cross-cultural research (Bleidorn et al., 2013) that found normative maturation was accelerated in countries in which work and family responsibilities were taken on sooner.

2c. **Arrested development.** People who experience a life event will show no change in personality, while those who did not will experience normative patterns of change. Evidence for arrested development comes from studies of incarceration in adolescence suggesting that adolescents who are incarcerated show minimal decreases in impulsivity relative to those who were not incarcerated (Bollich et al., 2017).

2d. **Retrogression.** Individuals who experience a life event will show non-normative declines in socially desirable personality traits, regardless of their baseline standing on a trait. Evidence for retrogression again comes from a longitudinal study of military training, in which individuals who joined the military decreased in Agreeableness after training and such decreases persisted even after leaving the military (Jackson et al., 2011).



We are atheoretical about our predictions for differences among people who did or did not experience different life events. We expect differences to be different across life events – that is, parenthood may change personality differently from beginning one's first job and retirement might impact personality differently than losing a child. Moreover, we do not expect our results to align with Specht et al.'s (2011) findings (see the table and summary below) using two waves of personality and life event data using the same sample because we expect that the propensity score matching group will partially explain some socialization effects they observed.

Rationale (Summary of Specht et al. (2011)):

- **Extraversion**: people who married, moved in with a partner, and left the parental home became less Extraverted
- **Agreeableness**: people who divorced or separated from their partner became more Agreeable
- **Conscientiousness**: People who get divorced or start their first jobs become more Conscientious, while people who have a child or retire become less Con-

scientious. Furthermore, women who lost a spouse became less Conscientious, while men who lost a spouse became more Conscientious

- **Emotional Stability**: women, but not men, became more Emotionally Stable after leaving their parents' home
- **Openness**: people who married or left the parental home became less open

	Extraversion	Agreeableness	Conscientiousness	Emotional Stability	Openness
Event	Slope	Slope	Slope	Slope	Slope
Marriage	↓				↓
Moved in with partner	↓				
Divorce		↑	↑		
Separation from partner		↑			
Death of a spouse					
Leaving parental home	↓				↓
Child leaves home					
Birth of a child			↓		
Death of a parent					
Unemployment					
Retirement			↓		
First Job			↑		
Sex x Marriage					
Sex x Leaving Parental Home				↓	
Sex x Separation from Partner					↑
Sex x Death of Spouse			↓		

Methods

Our project focuses on two groups in our sample: (a) people who experienced a life event and (b) people who did not. For each life event, we outline the item text used to gather information on life events in each wave of the GSOEP.

Category	Item	Item Text
Life Event	NewPart	Got Together with a New Partner
Life Event	Married	Married
Life Event	MoveIn	Moved In Together
Life Event	ChldBrth	Child Born
Life Event	ChldMvIn	Child Moved In
Life Event	ChldMvOut	Child Moved Out
Life Event	SepPart	Separated From Partner
Life Event	Divorce	Divorced
Life Event	PartDied	Partner Died
Life Event	DadDied	Father Died
Life Event	MomDied	Mother Died
Life Event	ChldDied	Child Died
Life Event	HHMDied	person living in hh died
Life Event	CHHComp	Other HH Comp Change
Life Event	NoCHHComp	No Change In HH Composition
Life Event	Unemploy	Registered Unemployed In Prev. Yr
Life Event	Retire	Retired In Prev. Yr
Life Event	FrstJob	Type Of Job Change
Life Event	LeftPar	Why Changed- Left Parents' House

We are studying the Big 5: Extraversion, Agreeableness, Conscientiousness, Emotional Stability (Neuroticism), and Openness to Experience. They were defined as following:

Category	Item	Item Text
Big 5	BF_C1	Thorough Worker
Big 5	BF_E1	Am communicative
Big 5	BF_A1	Am sometimes too coarse with others
Big 5	BF_O1	Am original
Big 5	BF_N1	Worry a lot
Big 5	BF_A2	Able to forgive
Big 5	BF_C2	Tend to be lazy
Big 5	BF_E2	Am sociable
Big 5	BF_O2	Value artistic experiences
Big 5	BF_N2	Somewhat nervous
Big 5	BF_C3	Carry out tasks efficiently
Big 5	BF_E3	Reserved
Big 5	BF_A3	Friendly with others
Big 5	BF_O3	Have a lively imagination
Big 5	BF_N3	Deal well with stress

Planned sample

This project uses the German Socioeconomic Panel Study (GSOEP) data. These data were collected by the German Institute of Economic Research (DIW Berlin) and are available, through application, at https://www.diw.de/soep.

Participants were recruited from more than 11,000 households and data have been collected annually since 1985. The latest data release includes data up to 2015. On average, 20,000 individuals are sampled each year. More information on the GSOEP can be found at https://www.diw.de/en/diw_02.c.221178.en/about_soep.html#299799, but, in short, the GSOEP is a nationally representative sample of private German households. It is critical to note that the GSOEP samples households, not individuals, and the households consist of individuals “living in both the”old" and “new” federal states (the former West and East Germany), foreigners, and recent immigrants to Germany."

Sample size varies by year, ranging from approximately 10,000 (1989) to 31,000 (2013) (http://soep.readthedocs.io/en/latest/overview/sample.html). This provides 99% power to detect a zero-order correlation effect size of ~.06, two-tailed at alpha

.05.

		Age in 2005		
Life Event	Frequency (Total)	*M*	*SD*	% women
ChldBrth	1279 (19607)	30.90	7.55	0.53
ChldMvOut	2148 (19599)	48.39	8.03	0.55
DadDied	963 (19615)	43.60	9.92	0.53
Divorce	413 (19614)	40.65	8.45	0.58
FrstJob	523 (19616)	22.16	4.98	0.54
LeftPar	449 (19611)	23.77	7.54	0.49
Married	1081 (19614)	33.30	10.91	0.52
MomDied	1016 (19618)	50.26	10.07	0.51
MoveIn	1073 (19602)	31.07	10.82	0.53
ParDied	1834 (19609)	46.95	10.73	0.52
PartDied	439 (19620)	64.74	11.77	0.70
Retire	4819 (19582)	64.80	9.04	0.52
SepPart	1128 (19607)	35.55	11.37	0.57
Unemploy	2297 (19584)	37.58	12.86	0.52

Exclusion criteria

Participants whose data fall outside the possible range of responses for a given question will be excluded.

To be included, participants had to have at least one wave of measurement in each of the following:

- **Matching variables** (variables used for multiple imputation and propensity score matching/weighting)
- **Grouping variables** (variables used to track the presence or absence of at least one life event)
- **Outcome variables** (personality measures for growth curves)

Analysis plan

Confirmatory analyses

Our confirmatory analyses will be tested using a series of second-order latent growth curve models.

Predictor Variables:

- Wave (2005, 2009, 2013)
- Group (life event or no life event): moderator variable

Outcome Variable(s):
Personality measured using the BFI-S (John, Donahue, & Kentle, 1991; see also John, Naumann, & Soto, 2008, and Lang, Lüdtke, & Asendorpf, 2001, for further information on the scale, the German translation, and evidence for its reliability and validity). The short version was created by Gerlitz & Schupp (2005) and contains 15 items that participants responded to on a Likert-like scale from 1 (does not apply at all) to 7 (applies perfectly).

Selection Effects: We will remove selection effects using propensity score matching.
Normative Development: the latent slope term in the latent growth model indicates the overall change trajectory.
Socialization Effects: regressing the life event group term on the latent slope in the latent growth curve models indicates group differences in personality change.

(Note that as shown in the model syntax below, each of the above will be tested with a single model for each trait-event combination.)

Covariates: Based on past work (e.g. Specht et al., 2011; Soto et al., 2011), we will include age and gender as covariates for each model. We will also test whether age and gender moderate the effect of life events on personality and personality change.

We will implement the 2nd-order latent growth curve models using the `blavaan` package in `R`. We plan to use the default priors, which reflect defaults in the the JAGS software imported by the `blavaan` package (see below).

```
##              nu           alpha          lambda            beta
```

```
## "dnorm(0,1e-3)" "dnorm(0,1e-2)" "dnorm(0,1e-2)" "dnorm(0,1e-2)"
##          itheta            ipsi             rho           ibpsi
##  "dgamma(1,.5)"  "dgamma(1,.5)"    "dbeta(1,1)" "dwish(iden,3)"
##             tau           delta
##   "dnorm(0,.1)"  "dgamma(1,.5)"
```

Details of analysis

The analysis phase will consist of four main parts, with interim steps to link these together: multiple imputation, propensity score matching, tests of longitudinal measurement invariance, and tests of socialization effects using bayesian latent growth curve modeling.

First, we will use multiple imputation to impute missing data for the matching variables (see below). Before doing so, we first create composites of our matching variables prior to 2005. We do this because irregularities in survey construction and response due to a variety of factors. To ensure transparency, we conduct all analyses using the raw data imported directly from the raw .sav files obtained from the SOEP website, and all steps in creating the composites are documented in a spreadsheet containing the item lists, text, and scales. Moreover, all steps are documented in a separate .Rmd file attached with this preregistration. The composite matching variables will later be used in propensity score matching, which requires completely non-missing data. Multiple imputation will be conducted using the `mi` package in R. Because the `missing_data.frame()` function attempts to automatically detect the scale of measurement, we will manually set all variables with ranges of at least 3 to continuous. We will impute 10 data sets. Full details of the multiple imputation procedure are available in the supplementary R scripts to this pre-registration.

Matching Variables

Category	Item	Item Text
Procedural	PROC_SID	Never Changing Person ID
Demographics	Dem_DOB	Year of Birth
Demographics	Dem_Sex	Sex
Procedural	PROC_household	household ID
Activities	Act_Volunteer	perform volunteer work

(continued)

Category	Item	Item Text
Background	Bkgr_DadPres	Father Present
Background	Bkgr_DisabStat	Disability Status of Individual
Background	Bkgr_Edu	Type of tertiary degree
Background	Bkgr_MarStat	Marital Status of individual
Background	Bkgr_MomPres	Mother Present
Background	Bkgr_PGovIncome	Pre-Government Income
Background	Bkgr_UrbOrRur	Spatial category by BBSR
Demographics	Dem_Race	Race of individual
Financial	Fnc_GrossSlry	Gross Amount Of Wages, Salary Prev. Yr
Financial	Fnc_HouseAssist	Housing Assistance
Financial	Fnc_StudGrnt	Student Grant
Financial	Fnc_UnempBen	unemployment benefit
Household	HH_BrothPres	Brother Present
Household	HH_ClnHlp	Cleaning Or Household Help In Household
Household	HH_CndHouse	Condition Of House
Household	HH_ColTV	Household Has Color Television
Household	HH_NumPer	Number Of Persons In Household
Household	HH_NumPer15to18	Number of hh members age 15-18
Household	HH_NumPerBel14	Number of hh members age 0-14
Household	HH_SisPres	Sister Present
Health	Hlth_BMI	Body Mass Index
Health	Hlth_BodPain	Bodily Pain (NBS)
Health	Hlth_EmoRole	Role Emotional (NBS)
Health	Hlth_GenHlth	General Health (NBS)
Health	Hlth_HeightCM	Body Height in cm
Health	Hlth_HlthInsr	Type Of Health Insurance
Health	Hlth_MntlHlth	Mental Health (NBS)
Health	Hlth_MntlSum	MCS: Summary Scale Mental (NBS)
Health	Hlth_NumDrVisits	Number of annual doctor visits

(continued)

Category	Item	Item Text
Health	Hlth_PhysFunc	Phyiscal Functioning (NBS)
Health	Hlth_PhysHlth	PCS: Summary Scale Physical
Health	Hlth_PhysProb	Accomplished Less Due To Physical Problems
Health	Hlth_PhysRole	Role Physical (NBS)
Health	Hlth_SocFunc	Social Functioning (NBS)
Health	Hlth_Vitality	Vitality (NBS)
Health	Hlth_WeightKG	Weight in kg
Psychological	Psych_LifeSat	Overall Life Satisfaction
Psychological	Psych_OthWorr	Other Worries
Psychological	Psych_SatFam	Satisfaction With Family Life
Psychological	Psych_SatHealth	Satisfaction with Health
Psychological	Psych_SatIncome	Satisfaction With Household Income
Psychological	Psych_SatSchool	Satisfaction With School Education and Vocational Retraining
Psychological	Psych_WorrCrm	Worried About Crime
Relationships	Rel_RelDad	Nature Of Relationship To Father
Relationships	Rel_RelMom	Nature Of Relationship To Mother
Social	Soc_SocGath	Attend Social Gatherings
Social	Soc_VisFam	Visit Family Members
Social	Soc_VisNghbr	Visit Neighbors, Friends

Second, we will use the multiply imputed data to calculate propensity scores for each of the mutliply imputed datasets for each life event separately. We will then use propensity score matching to try to equate our 2 groups in the sample. Because the sample size of the groups of people who experience specific life events are much smaller than the individuals who did not experience them, we choose to use propensity score matching, we begin by using "nearest neighbor" matching and a ratio of 4 to 1 and a caliper width of $.25\sigma$ (Guo & Fraser, 2015). The ratio will be updated iteratively for different life events if we are unable to achieve balance with our starting ratio of 4 to 1. Propensity score matching will be done using the

`matchit` package in R. After completing the propensity score weighting, we will examine the balance plots to test normality of the groups after matching.

Third, we will test for longitudinal invariance for each Big 5 factor. This will serve as the measurement model and the basis for all further models of the data. We have three waves of data with three indicators of each Big 5 trait at each wave, so we will construct a measurment model with 3 correlated latent variables (one for each wave), each consisting of 3 indicators. We impose strong invariance, meaning that we constrain the factor loadings and intercepts but not the error variances to be equal across all the waves. When this holds, this means that we can assume each latent variable intercept is on the same scale, which allows us to interpret slopes as change in the larger latent trait. Furthermore, we allowed each of the three indicators to correlate across time points, to account for the correlated error structure.

Fourth, we will merge the matched datasets with the outcome (personality) data for use in growth curve models. Because of equal spacing between personality waves, we will use second-order latent growth curve models. We will use `blavaan` package in R, which allows us to to use bayesian estiamtion of the LGCM. Age in 2005 and gender will be included as a covariate in all models, and the "no life event" group will be considered the reference group. The basic model of the latent growth curve model with random slopes and intercept is as follows:

We will fit growth curve models for each of the predictors for each of the multiply imputed datasets. To test for nonlinearity, we will fit latent basis models, eliminating the constraint on the loading of the latent slope in 2009 and 2013. We will make model selections based on comparisons among marginal log-likelihoods. Once they are fitted, we will use Rubin's Rules to pool fixed effects and random effect variance terms. Pooled individual level random effects will be calculated by taking the average across the datasets for each person individually (see the .Rmd script of planned analyses for more details). Results across the matched and unmatched pooled models will be extracted using the R packages `dplyr`, `tidyr`, and `purrr`. Average fixed and random effects across models will be displayed using simple slopes for each group from the growth model above using the R package `ggplot2`.

```
growth.mod <- '
T1 =~ NA*T1_1 + (lambda1)*T1_1 + (lambda2)*T1_2 + (lambda3)*T1_3
T2 =~ NA*T2_1 + (lambda1)*T2_1 + (lambda2)*T2_2 + (lambda3)*T2_3
```

```
T3 =~ NA*T3_1 + (lambda1)*T3_1 + (lambda2)*T3_2 + (lambda3)*T3_3



###intercepts
T1_1 ~ (nu1)*1
T1_2 ~ (nu2)*1
T1_3 ~ (nu3)*1
T2_1 ~ (nu1)*1
T2_2 ~ (nu2)*1
T2_3 ~ (nu3)*1
T3_1 ~ (nu1)*1
T3_2 ~ (nu2)*1
T3_3 ~ (nu3)*1

####variances/covariances
# item 1
T1_1 ~~ T2_1
T1_1 ~~ T3_1
T2_1 ~~ T3_1

# item 2
T1_2 ~~ T2_2
T1_2 ~~ T3_2
T2_2 ~~ T3_2

# item 3
T1_3 ~~ T2_3
T1_3 ~~ T3_3
T2_3 ~~ T3_3

I =~ 1*T1 + 1*T2 + 1*T3
S =~ 0*T1 + 1*T2 + 2*T3

I ~~ I
S ~~ S
I ~~ S
I ~ 1
```

```
S ~ 1
T1 ~ 0*1
T2 ~ 0*1
T3 ~ 0*1

#model constraints
lambda1 == 3 - lambda2 - lambda3
(nu1) == 0 - (nu2) - (nu3)

'


null.mod <- '
I ~ age + sex.c
'


group.mod <- '
  I ~ le.group + age.c + sex.c
  S ~ le.group + age.c + sex.c
'
```

Existing data

Data collection is underway or complete but I have not yet looked at this data.

These data were collected by the German Institute of Economic Research (DIW Berlin). Data have been downloaded and cleaned, but I have not run any statistical tests on the data except to calculate the number of individuals in each group for each life event to ensure that we had adequate power.

This study will be presented in fulfillment of the first author's master thesis at Washington University in St. Louis.
