



## **The Effect of Family Relationship on Happiness**

Thanh Huong NGUYEN

M2 Ingénierie économique et de l'analyse de données

Academic year: 2021 – 2022

Professor: Aristide E. HOUNDETOUNGAN

# TABLE OF CONTENTS

Introduction .....	3
Data .....	4
Strategies and Specifications .....	6
1. Pooled regressions .....	6
2. Random effect regressions .....	7
3. Fixed-effect regressions .....	7
Results .....	7
1. Pooled regressions .....	7
2. Random effect regressions .....	9
3. Fixed-effect regressions .....	10
Conclusion .....	12
Bibliography .....	13
Appendix A .....	14
Appendix B .....	15

# INTRODUCTION

Family is one of the building blocks of society. Relationship with family like parents, siblings, spouses and children can have a strong impact on happiness. Having a closed and satisfied relationship with loved ones means that you always have people who love, cherish and support you. Such relations are essential for our physical and mental health. This paper studies and measures, with empirical data, the impact of having a good relationship with loved ones on people level of happiness.

Using data from the National Longitudinal Study of Adolescent Health from Wave 1 to Wave 4, I tried to estimate the effects of family with three different models. The first approach is the **pooled estimator** using logit and probit regression. The estimation yielded a very statistically significant coefficient. It is estimated that having a good relationship with loved ones **increases the odds to be happy by about 5%**. The second approach estimates the impact using **random effects models**. I obtained a close result to the first approach, where good relationships can **increase the odds to be happy by nearly 4 times**. Finally, the third approach uses a **fixed-effects model**, which controlled for time-invariant factor. The result indicated an **increase by 1 percentage point of the probability to be happy** for people who maintain a good relationship with their loved ones. These findings can help address the social, behavioral, and health risk factors that contribute to changes in mental health.

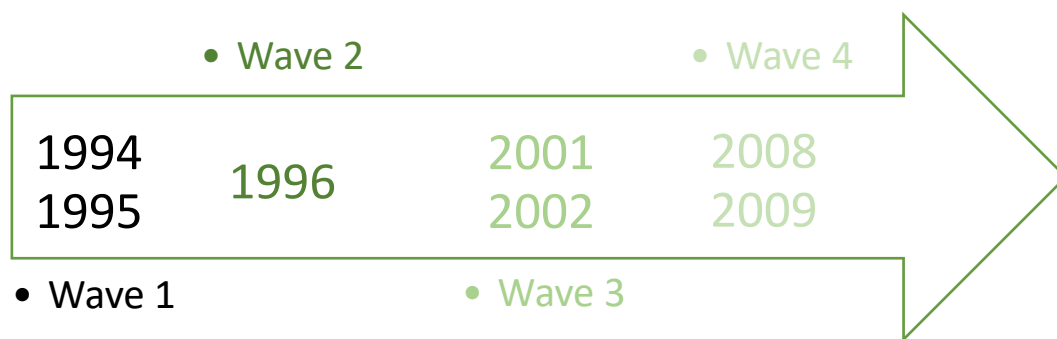
Several earlier research of this topics paved the way for my paper like the 2006 study of Haller & Hadler proved that social relations, occupational success, sociocultural orientations and participation are conducive to happiness and life satisfaction. Perelli-Harris et al. (2019) explored the relationship between cohabitation, marriage, and subjective well-being but they found no significant result. Another study focused on trust and marital relationship on subjective happiness by Riaz et al. (2016) resulted in fruitful conclusions.

This paper is organized as follows: **Section I** provides information on the survey and data manipulation. **Section II** introduces the empirical strategy, and results are discussed in **Section III**. Finally, **Section IV** provides concluding remarks.

## DATA

This study uses data on the status of relationship with loved ones to estimate the effects of relationship on mental health and emotions like happiness. Data is compiled from the National Longitudinal Study of Adolescent Health (Add Health) directed by Robert A. Hummer. This is a longitudinal study of a nationally representative sample of adolescents in grades 7-12 in the United States in 1994-95 that has been followed through adolescence and the transition to adulthood with four in-home interviews.

Add Health Wave I data collection took place between September 1994 and December 1995, when the individuals were 13 – 18 years old. Wave II was collected from April to August 1996 followed-up in-home interviews with adolescents from Wave I. Wave III data was collected from August 2001 to April 2002 with Wave I respondents who were then 18 to 26 years old. Wave IV in-home interviews were conducted in 2008 and 2009 when the original Wave I respondents were 24 to 32 years old (**Figure 1**). Another Wave V data collection took place from 2016 to 2018, but unfortunately, I was not able to acquire data of Wave 5.



*Figure 1. Evolution of Add Health survey*

Data from Wave 1 to Wave 4 are merge on respondents unique ID numbers. I also modify and create new variables for the regressions (parts of data manipulation were done using Stata because it was easier and quicker for me). I obtained a panel data of 6,500 adolescents, with some having observations for all 4 waves and some are not. Hence, the panel data structure is unbalanced.

I wanted to observe how relationship effect our well-being, or more specifically the state of happiness. But happiness is hard to define, it can be felt and perceived differently from person to person and nation to nation. In her 2007 book “The How of Happiness”, psychology researcher Sonja Lyubomirsky described happiness as “the

experience of joy, contentment, or positive well-being, combined with a sense that one's life is good, meaningful, and worthwhile". I decided to use this definition to measure happiness in my paper. Outcome variable *happy* is a binary variable with value 1 for feeling happy and 0 otherwise. Variable *happy* is calculated by combining different survey questions about the respondents' personalities and how they feel about their lives. The questions include : "You are physically fit?", "You have a lot to be proud of?", "You like yourself just the way you are?", "You feel like you are doing everything just about right?", "You feel socially accepted?", "You feel loved and wanted?", "How satisfied are you with your life as a whole?", "How often do you feel isolated from others?". Positive response from any of these questions would categorize the respondent as happy.

My variable of interest *family* is also a binary variable with value 1 for having a close and satisfied relationship with loved ones and 0 otherwise. Values were attributed by examining questions like: "How close do you feel to your loved ones?", "Are your loved ones warm and loving toward you?", "Are you satisfied with your relationship with your loved ones?". Loved ones here could be parents and siblings for younger individuals or spouses for married individuals. In some cases, they can be closest friends or dating partners or a mixture of people that are considered important by respondents.

I also controlled for individuals' characteristics by adding variable like age, age<sup>2</sup> and dummy for gender, race, marital status and health. Missing values are set to 0 for the regressions. The final dataset contains 21,324 observations and 14 variables.

year	Summary of Feeling happy		
	Mean	Std. Dev.	Freq.
1994	1	0	1
1995	.98425197	.12450873	6,477
1996	.98646679	.11555449	4,803
2001	.98520626	.12074367	3,515
2002	.98658718	.11507728	1,342
2007	1	0	46
2008	.99541559	.06755965	5,017
2009	1	0	47
Total	.98776355	.10994219	21,248

*Table 1. Frequency of variable happy by year of interview*

The dependent variable *happy* reports only 260 observations where respondents were feeling unhappy at some point in their lives (Table 2). Only 4% of all answers reports having no close relation with loved ones (Table 3).

Table 2. Happiness frequency

Feeling happy	Freq.	Percent
0	260	1.22
1	20,988	98.78
Total	21,248	100.00

Table 3. Close relations frequency

Close and satisfied relationship with loved ones	Freq.	Percent
0	879	4.12
1	20,445	95.88
Total	21,324	100.00

A more detailed report on the descriptive statistics is included in **Appendix A**.

## STRATEGIES AND SPECIFICATIONS

Since the dependent variable *happy* is binary, estimations using Ordinary Least Squared (OLS) would be highly bias. Therefore, in my paper, I will perform logistic regressions (Probit and Logit) with Maximum Likelihood (MV) estimations.

### 1. Pooled regressions

I used the **pooling** method to treat panel data as the normal cross-sectional data. This estimator is only convergent if the condition  $Cov(\alpha_i, x_{it}) = 0$  (the individual-specific effects are not correlated with the independent variables) is true. The effect is estimated using R command **plgm** with pooling method. My analysis is summarized by the following equation:

$$happy_{it} = \alpha + \beta^* family_{it} + \beta X'_{it} + \varepsilon_{it} \quad (1)$$

, where  $happy_{it}$  is a dummy for feeling happy of individual  $i$  in year of interview  $t$ . The coefficient of interest is  $\beta^*$ , associated to the dummy  $family_{it}$  that equals to 1 for having a close relationship with loved ones and 0 otherwise, which captures **the causal effect of having a good relationship with loved ones on the probability to be happy**.  $\alpha$  is the intercept for each individual at each point in time. I included  $X'_{it}$  as a set of control variables to control for individuals' characteristics that can affect happiness.

## 2. Random effect regressions

Here, I tried to estimate the same model, but the individual effects are treated as random effects. Meaning the variation across entities is assumed to be random and uncorrelated with the independent variables included in the model ( $Cov(\alpha_i, x_{it}) = 0$ ). I also took into account the panel structure of the data and estimated equation (1) using R command **glmer** with random intercepts (**1 | AID**), where the intercept varies with respondents. The integer scalar **nAGQ** is set to 0 for faster running time and to avoid error. This estimator is convergent and efficient in the case where the assumption is true.

## 3. Fixed-effect regressions

Fixed-effect estimator explore the relationship between independent variables and outcome variable within an individual. It implied correlation between individual-specific effects and independent variables ( $Cov(\alpha_i, x_{it}) \neq 0$ ). This method removes the effect of time-invariant characteristics (gender, etc.) so we can assess the net effect of the independent variables on the outcome variable.

$$happy_{it} = \alpha_i + \beta^*family_{it} + \beta X'_{it} + \varepsilon_{it} \quad (2)$$

, where  $\alpha_i$  ( $i = 1, \dots, n$ ) is the unknown intercept specific to each respondent. Fixed-effects does not work well with data for which within-cluster variation is minimal or for slow changing variables over time. I estimate equation (2) using **bife** without time-invariant variable (gender, race).

# RESULTS

## 1. Pooled regressions

**Table 4** and **Table 5** present the results from two pooled regressions. Both models give statistically significant *family* coefficients at 1% level. Log likelihood for the probit model is slightly smaller. Considering the logit model, the odds of being happy for people with a good relationship with their loved ones ( $family = 1$ ) over the odds of being happy for people without a good relationship with their loved ones ( $family = 0$ ) is  $exp^{1.639931} = 5,155$ . Or people with a good relationship with their loved ones have 415% higher odds than people who do not. Having a good relationship with loved ones increases the probability of being happy by 4,92% points (logit) or 5,24% points (probit). Results on marginal effects is reported in **Appendix B**.



*Table 4. Pooled logit regression.*

Logistic regression		Number of obs	=	21,324		
		Wald chi2(9)	=	354.55		
		Prob > chi2	=	0.0000		
Log pseudolikelihood = -1577.1357		Pseudo R2	=	0.0873		
(Std. Err. adjusted for 6,500 clusters in AID)						
happy	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
family	1.639931***	.1416594	11.58	0.000	1.362284	1.917579
female	-.2313973	.122237	-1.89	0.058	-.4709774	.0081828
black	.7546176***	.2430362	3.10	0.002	.2782754	1.23096
asian	-.1406447	.3135098	-0.45	0.654	-.7551126	.4738233
white	.2867244	.2114245	1.36	0.175	-.12766	.7011087
ms	.1804595	.2149966	0.84	0.401	-.240926	.6018451
health	1.396325***	.1442909	9.68	0.000	1.11352	1.67913
age	-.7370763***	.1291201	-5.71	0.000	-.9901471	-.4840055
age2	.0183747***	.003007	6.11	0.000	.0124811	.0242683
_cons	8.193013***	1.336289	6.13	0.000	5.573935	10.81209

Note: Robust standard errors, clustered at individual level. \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

*Table 5. Pooled probit regression.*

Probit regression		Number of obs	=	21,324		
		Wald chi2(9)	=	289.28		
		Prob > chi2	=	0.0000		
Log pseudolikelihood = -1577.0732		Pseudo R2	=	0.0873		
(Std. Err. adjusted for 6,500 clusters in AID)						
happy	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
family	.7428141***	.0674185	11.02	0.000	.6106764	.8749519
female	-.0900488	.0493231	-1.83	0.068	-.1867202	.0066226
black	.3003835***	.1000239	3.00	0.003	.1043402	.4964268
asian	-.073128	.1301998	-0.56	0.574	-.3283149	.182059
white	.1169381	.0886947	1.32	0.187	-.0569003	.2907766
ms	.055835	.0840418	0.66	0.506	-.1088839	.2205538
health	.604276***	.0654895	9.23	0.000	.475919	.732633
age	-.2923321***	.0495365	-5.90	0.000	-.3894218	-.1952423
age2	.0072659***	.0011491	6.32	0.000	.0050136	.0095181
_cons	3.621938***	.5131651	7.06	0.000	2.616153	4.627724

Note: Robust standard errors, clustered at individual level. \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .



## 2. Random effect regressions

**Table 6** and **Table 7** present the results from logit and probit regression for panel data. Both models give very statistically significant *family* coefficients. Variance of random effects is not null (1,218 and 0,1818), which means it was right to control for the variance within respondent. Positive coefficients for *family* show that people with a good relationship with their loved ones are more likely to be happy. Considering again the logit model, the odds of being happy for people with a good relationship with their loved ones over the odds of being happy for people without a good relationship with their loved ones is  $\exp^{1.537738} = 4,65$ . Or people with a good relationship with their loved ones have nearly 4 times higher odds than people who don't.

**Table 6.** Panel logit regression with random effects

```
Generalized linear mixed model fit by maximum likelihood (Adaptive Gauss-Hermite Quadrature, nAGQ = 0)
[glmerMod]
Family: binomial ( logit )
Formula: happy ~ family + female + age + age2 + black + asian + white + ms + health + (1 | AID)
Data: data
Control: glmerControl(optimizer = "bobyqa")

      AIC      BIC    logLik deviance df.resid
 3145.0   3232.7  -1561.5   3123.0   21313

Scaled residuals:
    Min       1Q   Median       3Q      Max
-18.6001   0.0827   0.1055   0.1229   0.8574

Random effects:
 Groups Name      Variance Std.Dev.
 AID      (Intercept) 1.218    1.103
Number of obs: 21324, groups: AID, 6500

Fixed effects:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  8.436872   1.301191   6.484 8.93e-11 ***
family       1.537738   0.152711  10.070 < 2e-16 ***
female      -0.220119   0.121082   -1.818 0.069073 .
age         -0.740656   0.125793   -5.888 3.91e-09 ***
age2         0.018477   0.002972    6.217 5.05e-10 ***
black        0.725371   0.217560    3.334 0.000856 ***
asian       -0.141769   0.286988   -0.494 0.621315
white        0.284775   0.185501    1.535 0.124741
ms           0.182662   0.224230    0.815 0.415291
health       1.322818   0.147054    8.995 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
      (Intr) family female age   age2   black  asian  white  ms
family -0.058
female -0.069 0.063
age    -0.976 -0.022 0.000
age2   0.957 0.007 0.005 -0.992
black  -0.097 -0.004 -0.024 -0.002 -0.002
asian  -0.059 0.000 0.009 -0.006 0.004 0.387
white  -0.123 -0.026 -0.005 0.005 -0.004 0.699 0.460
ms     -0.012 0.011 -0.084 0.046 -0.100 0.045 0.023 -0.012
health -0.018 -0.073 0.059 -0.070 0.072 -0.002 -0.044 -0.030 0.018
```

*Table 7. Panel probit regression with random effects*

```

Generalized linear mixed model fit by maximum likelihood (Adaptive Gauss-Hermite Quadrature,
nAGQ = 0) [glmerMod]
Family: binomial (probit)
Formula: happy ~ family + female + age + age2 + black + asian + white +
ms + health + (1 | AID)
Data: data
Control: glmerControl(optimizer = "bobyqa")

      AIC      BIC    logLik deviance df.resid
3151.7   3239.4  -1564.9   3129.7    21313

Scaled residuals:
      Min       1Q   Median       3Q      Max
-18.8410   0.0771   0.0992   0.1178   0.6947

Random effects:
Groups Name      Variance Std.Dev.
AID (Intercept) 0.1818   0.4263
Number of obs: 21324, groups: AID, 6500

Fixed effects:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  3.697047   0.531122  6.961 3.38e-12 ***
family       0.719156   0.072447  9.927 < 2e-16 ***
female      -0.087618   0.050422  -1.738 0.08226 .
age         -0.291019   0.051115  -5.693 1.25e-08 ***
age2         0.007244   0.001200   6.036 1.58e-09 ***
black        0.296269   0.091772   3.228 0.00125 **
asian       -0.076889   0.123354  -0.623 0.53307
white        0.115667   0.079984   1.446 0.14814
ms           0.057617   0.091986   0.626 0.53108
health       0.594280   0.067597   8.791 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
      (Intr) family female age    age2    black    asian    white    ms
family -0.083
female -0.071 0.053
age    -0.971 -0.021 0.002
age2    0.951 0.011 0.003 -0.992
black  -0.103 0.005 -0.020 -0.006 0.003
asian  -0.065 0.001 0.012 -0.004 0.003 0.398
white  -0.127 -0.020 0.001 0.000 0.001 0.719 0.466
ms     -0.018 0.002 -0.085 0.055 -0.112 0.047 0.021 -0.013
health -0.018 -0.055 0.055 -0.089 0.094 0.003 -0.041 -0.032 0.015

```

### 3. Fixed-effect regressions

Dependent variable *happy* is high skew and does not change for most of the respondents (20,324 observations were dropped because of all positive outcomes). The regression used only 1000 observations (286 respondents). I also corrected analytical bias (**bias\_corr**) to reduce the incidental parameter bias problem. Results are reported in **Table 8** and **Table 9**.

*Table 8. Fixed-effects logit model with corrected bias*

```

binomial - logit link

happy ~ family + age + age2 + ms + health | AID

Estimates:
      Estimate Std. error z value Pr(> |z|)
family  0.74425    0.26432   2.816  0.00487 **
age     -0.72437    0.15765  -4.595  4.33e-06 ***
age2     0.01806    0.00360   5.015  5.29e-07 ***
ms       0.24682    0.31971   0.772  0.44009
health   0.50201    0.24456   2.053  0.04010 *
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

residual deviance= 1087.03,
null deviance= 1250.71,
nT= 1000, N= 286

( 20324 observation(s) deleted due to perfect classification )

Number of Fisher Scoring Iterations: 5

Average individual fixed effect= 6.452

```

*Table 9. Fixed-effects probit model with corrected bias*

```

binomial - probit link

happy ~ family + age + age2 + ms + health | AID

Estimates:
      Estimate Std. error z value Pr(> |z|)
family  0.49424    0.15886   3.111  0.00186 **
age     -0.41246    0.09045  -4.560  5.11e-06 ***
age2     0.01030    0.00205   5.024  5.07e-07 ***
ms       0.13075    0.18643   0.701  0.48309
health   0.31730    0.14531   2.184  0.02899 *
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

residual deviance= 1085.53,
null deviance= 1250.71,
nT= 1000, N= 286

( 20324 observation(s) deleted due to perfect classification )

Number of Fisher Scoring Iterations: 6

Average individual fixed effect= 3.614

```

*Table 10. Average Partial Effects*

	Probit		Logit	
	Corrected	Uncorrected	Corrected	Uncorrected
<i>family</i>	0.00989 *** (0.00268)	0.01096 *** (0.00272)	0.00939 *** (0.00272)	0.01084 *** (0.00277)
<i>age</i>	-0.00763 *** (0.00128)	-0.00830 *** (0.00143)	-0.00831 *** (0.00135)	-0.00869 *** (0.00153)
<i>age2</i>	0.00019 *** (0.00003)	0.00021 *** (0.00003)	0.00021 *** (0.00003)	0.00022 *** (0.00004)
<i>ms</i>	0.00237 (0.00235)	0.00257 (0.00243)	0.00274 *** (0.00242)	0.00293 *** (0.00250)
<i>health</i>	0.00611 *** (0.00227)	0.00664 *** (0.00233)	0.00606 *** (0.00232)	0.00664 *** (0.00240)

Note: \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

**Table 10** shows the estimates of the average partial effects (APEs) and the corresponding standard errors (in parentheses). I observe that the bias-corrected and uncorrected estimates are almost identical for the probit model and slightly different for the logit model. Overall, having a close relationship with loved ones increases the probability to be happy by 1% percentage points.

## CONCLUSION

Using the Add Healthy survey data, I explored the link between relationship with loved ones and the state of happiness. The **pooling method** resulted in **an increase of 5%** of the odds to be happy for people having good relationship with loved ones. Similarly, the **random effects model** gives a very statistically significant result. By having satisfying relationships, the odds to be happy is **increased by nearly 4%**. Lastly, the **fixed-effects model** provides estimates of the average partial effect. Both bias-corrected logit and probit models shows a statistically significant estimate of **an increase by 1% point**. By controlling for random and fixed effects, the log odds of being happy reduces significantly. If the correlation assumption between individual-specific effect and covariates is true, fixed-effects estimator is convergent and not bias. Otherwise, random effects estimator is convergent and also efficient.

## BIBLIOGRAPHY

4. Lyubomirsky, S. (2010). *The how of happiness*. Piatkus Books.
5. Haller, M., Hadler, M. How Social Relations and Structures can Produce Happiness and Unhappiness: An International Comparative Analysis. *Soc Indic Res* 75, 169–216 (2006).
6. Perelli-Harris, B., Hoherz, S., Lappegard, T., & Evans, A. (2019). Mind the “Happiness” Gap: The Relationship between Cohabitation, Marriage, and Subjective Well-Being in the United Kingdom, Australia, Germany, and Norway. *Demography*, 56(4), 1219–1246.
7. Riaz, W., Riaz, M. N., & Zohra, N. I. (2016). Mutual Trust, Marital Relationship Satisfaction and Subjective Feelings of Happiness in Intimate Relationships. *Journal of Asian Development Studies*, 5(2), 41–49.

## APPENDIX A

Variable		Mean	Std. Dev.	Min	Max	Observations
AID	overall	9.42e+07	6486364	5.71e+07	9.97e+07	N = 21324
	between		6551601	5.71e+07	9.97e+07	n = 6500
	within		0	9.42e+07	9.42e+07	T-bar = 3.28062
birthyr	overall	1979.087	1.738871	1974	1983	N = 21324
	between		1.772885	1974	1983	n = 6500
	within		0	1979.087	1979.087	T-bar = 3.28062
year	overall	1999.779	5.192537	1994	2009	N = 21324
	between		1.790998	1995	2002	n = 6500
	within		5.002073	1992.779	2008.779	T-bar = 3.28062
white	overall	.668777	.4706641	0	1	N = 21324
	between		.4735891	0	1	n = 6500
	within		0	.668777	.668777	T-bar = 3.28062
black	overall	.2447946	.4299754	0	1	N = 21324
	between		.4324224	0	1	n = 6500
	within		0	.2447946	.2447946	T-bar = 3.28062
native	overall	.0370006	.1887676	0	1	N = 21324
	between		.1866872	0	1	n = 6500
	within		0	.0370006	.0370006	T-bar = 3.28062
asian	overall	.0399081	.1957479	0	1	N = 21324
	between		.1991935	0	1	n = 6500
	within		0	.0399081	.0399081	T-bar = 3.28062
ms	overall	.1643688	.3706186	0	1	N = 21324
	between		.214028	0	1	n = 6500
	within		.3044045	-.5856312	.9143688	T-bar = 3.28062
family	overall	.9587788	.198806	0	1	N = 21324
	between		.1400162	0	1	n = 6500
	within		.1584361	.2087788	1.708779	T-bar = 3.28062
happy	overall	.9877636	.1099422	0	1	N = 21248
	between		.0718659	0	1	n = 6495
	within		.0877223	.2377636	1.737764	T-bar = 3.27144
health	overall	.9321199	.2515459	0	1	N = 21317
	between		.1724564	0	1	n = 6499
	within		.1880342	.1821199	1.68212	T-bar = 3.28004
age	overall	20.69185	5.521061	12	34	N = 21324
	between		2.588192	13	27.66667	n = 6500
	within		5.002073	13.69185	29.69185	T-bar = 3.28062
female	overall	.5278559	.4992351	0	1	N = 21324
	between		.4997824	0	1	n = 6500
	within		0	.5278559	.5278559	T-bar = 3.28062
age2	overall	458.6333	246.8121	144	1156	N = 21324
	between		112.7404	169	798.5	n = 6500
	within		225.0151	101.1333	912.9667	T-bar = 3.28062



