Joint Modeling of Longitudinal Outcome and Time to Drop out – An Application Using Complex Survey Data

Punam Pahwa[1], Chandima Karunanayake[2]
[1]Community Health and Epidemiology, University of Saskatchewan, 103 Hospital Drive, Saskatoon, SK, S7N OW8, Canada
[2]Canadian Centre for Health and safety in Agriculture, University of Saskatchewan, 103 Hospital Drive, Saskatoon, SK, S7N OW8, Canada

Abstract

Statistics Canada National Population Health Survey (NPHS) data were used to trace longitudinal changes in the mental health of immigrant women. In the NPHS a member from each household was randomly chosen to participate in this longitudinal survey. We explored the moderating affects of different risk factors on immigrant women's mental distress, which was categorized as a dichotomous variable: no/low distress vs. moderate/high distress. Regression coefficients were estimated by using the weighted generalized estimating equations approach which accounts for within-subject correlation and variance estimation was conducted by taking into account the complexities of multi-stage design using bootstrap weights computed by Statistics Canada. An alternative approach based on the joint model for a longitudinal outcome and the time to dropout is under process.

KEY WORDS: Generalized Estimating Equations, Complex Survey, Joint Modeling

INTRODUCTION:

Many longitudinal studies generate both longitudinal (repeated measurements) and survival (time to an event) data. Like standard longitudinal studies, in longitudinal complex surveys, repeated measurements on a quantitative (continuous or discrete) response, and information on time-dependent and time-independent covariates is collected on each participant[1]. Additional information, either on time to drop out or time to occurance of any other event of interest and variables related to such data increases researcher's interest to investigate the interrelationships between longitudinal and censored time-to-event data[1]. Investigating such interrelationships complicates the analysis. A challenging problem is a joint modeling of these data to explicate the association between these two types of responses. By joint modeling the event time, the analysis of longitudinal measurements is adjusted to allow for non-ignorable missing data due to informative dropout, which cannot be appropriately handled by the standard linear mixed effects models alone. According to Shen and Weissfeld[2], it is impossible to distinguish between missing at random (MAR) and missing not at random (MNAR) based on the observed data. A safe strategy is to treat the data as MNAR, which is non-ignorable.

In the recent years, several approaches for joint modeling of repeated measurements and time to event data have been researched and applied to analyze real life data. Joint analysis of longitudinal measurements and survival data has received much attention in recent years. By modeling the event time, the analysis of longitudinal measurements is adjusted to allow for non-ignorable missing data due to informative dropout, which cannot be appropriately handled by the standard linear mixed effects models alone. In medical research, there are several examples where one can have repeated measurements and censored time-to-event data. Several statistical methods exist for analyzing such data separately, such as random/mixed effects model for longitudinal continuous outcome based on maximum likelihood estimation[3] , marginal or transitional models for longitudinal discrete outcome based on generalized estimating equations[3], and survival model (for example: Cox regression model) for time to event or variance corrected models for repeated event data[4]. However, the separate use of these models may be inappropriate, specially when the longitudinal response is correlated with study dropout[5]. Several approaches for the joint modeling have been proposed by various researchers in the recent years. Schulchter[6] and DeGruttola and Tu[6] used shared random-effects approaches to model log-survival time and a biomarker by using multivariate distribution in order to construct a joint model for a continuous and an event time process. Tsiatis et al[7]; and Wulfson and Tsiatis[8] reported that unbiased statistical inferences are more likely to be obtained via joint model. Wu and Carroll[9] and Hogan and Laird[10,11] discussed the difficulties for making inference on the longitudinal process when there is an additional time-to-event data which may induce an informative censoring. For example, subjects with serious HIV disease may tend to withdraw from the study earlier compared to the healthier subjects, leading to fewer CD4 measurements , and to have sharper rates of CD4 decline.

Henderson et al.[12] proposed a flexible joint model that avoids specifying the class variable, yet allows a very broad range of dependencies between the longitudinal responses and the survival endpoints. Guo and Carlin[13] investigated the approach proposed by Hendrson et al. by using standard computer packages.

The primary focus of approaches discussed above was to make inferences about the event time process , however and lacked prediction concept of data application. Bowman and Manatunga[14] made inferences about the joint process and discussed the prediction aspect, which is an important part of data application.

Like standard longitudinal studies, it is common to have repeated measurements on the response variables and drop-outs in longitudinal complex surveys. In this manuscript our interest is to (i) understand longitudinal patterns of changes of mental health by modeling the repeated measurements: (a) using the generalized estimating equations approach for dichotomous outcome and (b) maximum likelihood estimation for continuous outcome; (ii) investigate the variables affecting the time to drop out variable by using Cox's proportional hazard model and (iii) jointly model the relationship between mental health distress score and time to drop-out based on likelihood methods.

STATISTICAL METHDOLOGY:

Objective (i-a). Longitudinal Model based on dichotomous outcome:

A logistic regression model is assumed for the marginal mean of $Y_i(t)$,

$$p_i(t) = E[Y_i(t) \mid X_i(t)] = pr\{Y_i(t)\} = \frac{\exp\{X_i(t)\beta\}}{1+\exp\{X_i(t)\beta\}}$$

where β is the vector of logistic regression parameters, $\boldsymbol{X_i(t)}$ is a design matrix. SAS procedure PROC GENMOD was used to fit the binary logistic regression model.

Objective (i-b). Longitudinal model based on continuous outcome:

Let Y_{ij} be the observed value for the response variable for i^{th} subject at j^{th} time point and is modeled as:

$$y_{ij} = X_{1i}^{'}(s)\beta_1 + d_{1i}^{'}(s)u_i + \varepsilon_{ij}$$

$X_{1i}^{'}$ is the design matrix for i^{th} subject and $d_{1i}^{'}u_i$ incorporates subject-specific random effects into the above model and ε_{ij} are mutually independent random error terms and $\varepsilon_{ij} \sim N(0,\sigma_\varepsilon^2)$.

The u_i is a vector of random effects corresponding to the subject-specific explanatory variables ($d_{1i}(s)$), which may be a subset of design matrix $X_{1i}^{'}(s)$. SAS procedure PROC MIXED was used to fit the random effects model.

Objective (ii). Survival model for time to dropout data:

Cox proportional hazard model (also known as semi-parametric model) was sued to model the survival data, i.e time to drouput.

$$\lambda_i(t) = \lambda_0(t)\exp\{X_{2i}^{'}(t)\beta_2 + W_{2i}(t)\}$$

where baseline hazard function $\lambda_o(t)$ has an arbitrary form and $W_{2i}(t)$ include subject-specific random effects also known as frailty. SAS procedure PROC PHREG was used to fit the survival model for time to dropout data.

Objective (iii) - Joint modeling of mental of longitudinal mental health distress score and time to dropout data:

In order to address this objective, i.e characterizing association among the longitudinal and time-to-event processes and covariates, is to represent the relationship between T_i, $X_i(u)$ and Z_i by a proportional hazard model. We used joint modeling approach and macro developed by Guo and Carlin[13] to achieve this objective. According to Guo and Carlin[13], association between the longitudinal and survival process can arise in two ways: (i) through common explanatory variables or (ii) through stochastic dependence between subject-specific random effects. SAS procedure PROC NLMIXED was used to conduct the joint modeling.

Estimation of regression coefficients:

The SAS procedures PROC GENMOD, PROC MIXED, PROC PHREG and PROC NLMIXED[15] were used for binary logistic; random effects; survival model and joint modeling approaches respectively. These procedures were used to fit the multivariable models in order to determine the significant predictors of mental distress of immigrant women.. The longitudinal weight variable computed by the methodologists of Statistics Canada was used in the WEIGHT statement of SAS procedures.

Variance estimation:
Robust Variance estimation based on the GEEs and not accounting for the design:

Robust variance estimation in GENMOD is based on Zeger and Liang's method[16,17] which accounts only for the within-subject dependencies due to the repeated measurements over time. The variance estimation was based on the formula given by Liang and Zeger [16,17].

Survey Bootstrap for Variance Estimation accounting for the design complexities:

Statistics Canada releases design information for variance estimation only in the form of bootstrap weights: cross-sectional weights and longitudinal weights (adjusted for non-response) that have been created from taking numerous bootstrap samples of primary sampling units from the original sample. Computation of replicate survey weights is done by the methodologists of Statistics Canada[18] . A Bootstrap replication method was used that made appropriate use of these longitudinal bootstrap weights for the variance estimation of regression estimates. To account for the complexities of the multi-stage stratified clustered design the BOOTVAR program which was developed by Statistics Canada was used for the variance estimation.

Application to Longitudinal NPHS Data

Description of Data Set:

The Canadian NPHS was launched in 1994-1995[1]. The longitudinal sample data consist of 17,276 participants and this group of individuals will be surveyed every two years in future until 2014. Details of this survey are given elsewhere[19] .

Study Population:

Our aim was to compare mental health of immigrant women with those who were Canadian born. The study population consist of women 15 years and older.

The NPHS includes a set of questions designed to determine/investigate the mental health of NPHS participants. In this report, we used mental distress as a measure of mental health. The mental distress variable was derived from a set of questions designed by Kessler et al[20]. Our study population consists of females 15 years and older.

Dependent Variable:

Distress, an ordinal outcome variable was examined using a six-item scale that assessed feelings of i) sadness, ii) nervousness, iii) restlessness, iv) hopelessness , v) worthlessness and vi) the feeling that everything was an effort within the previous month. The variable "distress scale", is based on the work of Kessler and Morczek[20] and was derived from the Composite International Diagnostic Interview. Scores on the distress scales ranged from 0 (no distress) to 24 (highly distressed). The distribution of this distress scale was highly skewed, hence the outcome variable was categorized into two categories: i) no or low distress : 0-5 and ii) moderate/high distress: 6-24.

Independent variables:

Mental health is an interplay among several factors, such as: demographic; socio-economic, social-support, health related, time of study and interactions between them. In this report the following variables were considered as independent variables:

Main risk factors of interest: **Immigrant Status**: All NPHS participants who were not Canadian citizen by birth were defined as immigrants.
Ethnicity: Immigrant's origin is grouped into seven categories according to their country of birth: British, Eastern European, Western European, Chinese, South Asian, Black and other. This categorization was done in order to distinguish groups with cultural differences.

Demographic variables consist of age, marital status, location of residence, geographical area, and length of stay in Canada. **Age** was used as a time-dependent variable with four categories: 15-24 yrs, 25-54 yrs, 55-69 yrs and 70 yrs and older (reference category: 70 yrs and older).

Social Role**:** This variable had five categories**:** single parent, not employed; single parent employed; partnered with children, not employed; partnered with children, employed; women without children, partnered or single (reference category).

Socio-economic status variables consist of education and income. Education was a dichotomous variable with two categories: education received less than or equal to12 years and education received greater than 12 years. Income was divided into three categories based on the work of Wang and EI-Gebaly[21].

Social Support variables consist of a social involvement score, which was divided into three categories: low (0-1); moderate (2-4); and high (5-8). This score was based on two questions : frequency of participation in organizations and frequency of attending religious services.

Life-style variables consist of participant's personal smoking history and household smoking status. Personal Smoking history was divided into three categories, non-smokers, ex-smokers and current smoker. Household smoking status was a dichotomous variable indicating presence or absence of smokers within a household.

Health related variable consists of a self-perceived general health status, which had five categories: poor, fair, good, vary good, and excellent (reference category: excellent).

Four dummy variables for 'Cycle' was used to study the effect of time on mental distress.

Results:

Conclusions - Based on Longitudinal Binary Outcome:

Significant Main Effects: [OR & 95% C.I]

Age:

15-24 – [2.34, (1.85, 2.97)]
25-54 – [1.86, (1.60, 2.17)]
55-69 – [1.15, (0.98, 1.34)]

Social Role: Single parent, not employed at higher risk of moderate/high distress [1.59; (1.23,2.06)]

Location of residence: Rural females were at lower risk of having moderate/high distress: [0.87 (0.78, 0.96)]
Quebec women were at higher risk– [1.56 (1.36,1.80)]

Stay in Canada: Less than 2 yrs – [0.87, (0.49,1.55)]
>2 and <20 yrs - [1.46 (1.22,1.76)]

Income: Low – [1.44 (1.24,1.67)]
Middle – [1.09 (0.97,1.23)]

Self-perceived General Health status:

Poor: 14.99 (11.81, 19.04)
Fair: 6.05 (5.18, 7.06)
Good: 2.89 (2.55, 3.28)
Very good: 1.53 (1.36, 1.72)

Significant Interactions:

- Smoking * Ethnicity (these results are shown in figure 1.)
- Education*Ethnicity
- Social Involvement Score * Ethnicity
- Household Smoking*Ethnicity

Conclusions- Based on Survival Analysis

Younger women were more likely to drop out. The risk of dropping out decreases with age.

Chinese, South Asians and Black women were more likely to drop out compared to British women.

Eastern and Western European women were less likely to drop out compared to British women.

Rural women were less likely to drop out compared to Urban women.

Women from British Columbia were more likely to drop out compared to women from Ontario.

Women with length of stay in Canada less than 2 years is more likely to drop out compared to those who have length of stay more than 20 years.

Women with education level less than 12 years is more likely to drop out.

Women with low or middle income were more likely to drop out.

Current smokers and ex-smokers were less likely to drop out compared to non-smokers.

Women exposed to household smoking were more likely to drop out.

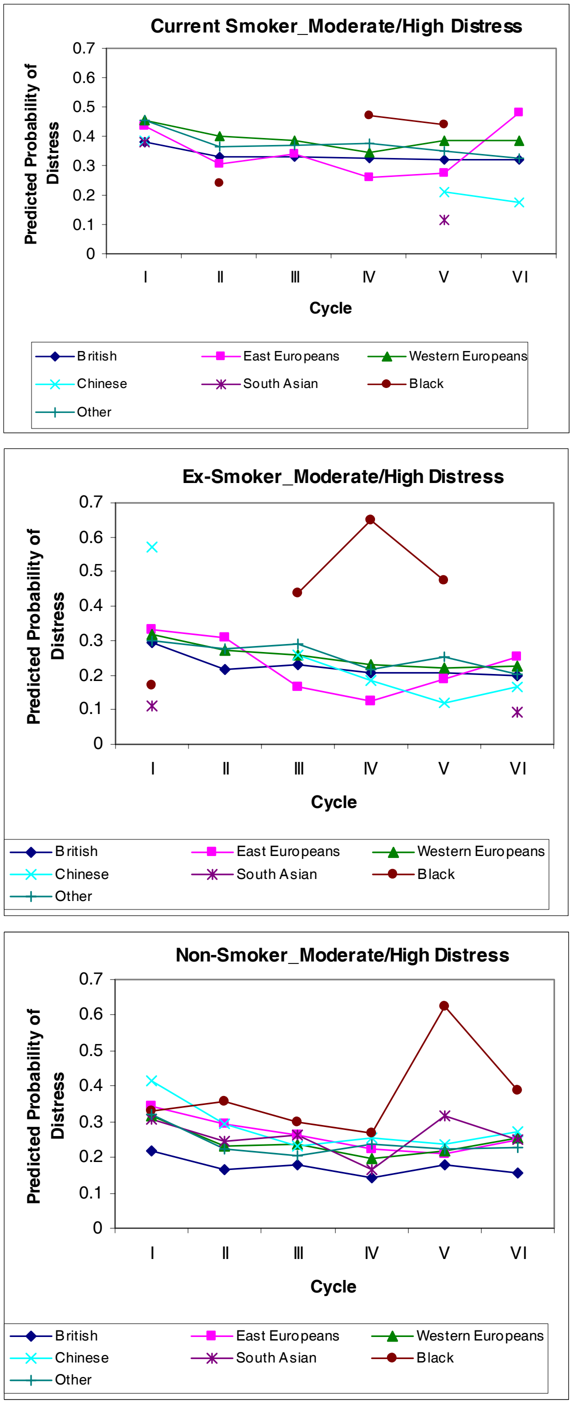

Figure 1. Predicted probability of mental distress scores stratified by smoking status and ethnicity status

Discussion:

In this manuscript we investigated longitudinal patterns of changes of mental health by modeling the repeated measurements: using the generalized estimating equations approach for dichotomous outcome and maximum likelihood estimation techniques for continuous outcome. We used the Cox's proportional hazard model to determine the variables which affect the time to drop out variable. We attempted to jointly model the relationship between mental health distress score and risk of drop-out dependent on the ethnicity of the study participant adjusted for other covariates. The application of joint model is important when we have longitudinal repeated measurements and time to event data for accurate inference for longitudinal responses while adjusting for outcome-dependent study dropout. We used PROC NLMIXED to fit the joint model based on maximum likelihood estimation, which allows random effects.

In the last few years, research on joint modeling of repeated measurements and time to event data continues to present new challenges. Some of the theoretical properties of these methods are still unresolved. We faced several challenges while conducting the analyses described in this manuscript. Some of the problems are described below.

Problems faced with bootstrap estimation:

While conducting the logistic regression analysis, we used a macro developed by Statistics Canada to estimate bootstrap variance estimates of regression parameters and some of the errors we obtained were: (i) error in computing the variance function; (ii) error in parameter estimate covariance computation and (iii) the generalized Hessian matrix is not positive definite.

Problems with Joint Modeling:

We used PROC NLMIXED macro developed by Guo and Carlin[13] to jointly model the longitudinal measurements and time-to-event data. We encountered an error 'Execution error for observation 54'. It was not possible to fix this error because we used Remote Data Access Services of Statistics Canada to conduct our research. After successful execution of SAS syntax on the dummy data provided by Statistics Canada, we sent our syntax to RDA office to run it on the original master data file and they provided us with the results.

Because of the problems/challenges faced during the joint modeling, we were not able to compare the results of these models with those of pattern-mixture model (presented in Table 1), a method commonly used to account for missing data.

REFERENCES:

1. Tsiatis A.A. and Davidian M. (2004). Joint modeling of longitudinal and time-to-event data: An overview. Statistica Sinica **14**(2004), 809-834
2. Shen and Weissfeld. (2005)
3. Diggle P.J., Liang K-Y and Zeger S.L. (1994). Analysis of longitudinal data, Oxford University Press, Oxford.
4. Therneau T.M and Grambsch P.M. (2001). Modeling Survival Data: Extending the Cox Model
5. Schluchter, M. (1992). Methods for the analysis of informatively censored longitudinal data. Statistics in Medicine, vol. 11:1861-1870.
6. DeGruttola, V. and Tu, X.M. (1994). Modeling progression of CD4-lymphocyte count and its relationship to survival time. Biometrics, vol. 50:1003-1014.
7. Tsiatis et al (1995)
8. Wulfson, M.S. and Tsiatis, A.A. (1997). A joint mmodel for survival and longitudinal data measured with error. Biometrics, Vol. 53:330-339.
9. Wu, M.C and Carroll, R.J. (1988). Estimation and comparison of changes in the presence of informative right censoing by modeling the censoring process. Biometrics, Vol. 44: 175-188.
10. Hogan, J.W. and Laird, N.W. (1997a). Mixture models for the joint distribution of repeated measures and event times. Statistics in Medicine, Vol. 16: 239-257.
11. Hogan, J.W and Laird, N.W. (1997b). Model-based approaches to analyzing incomplete longitudinal and failure time data. Biostatistics, Vol 1(4): 465-480.
12. Henderson R, Diggle P, Dobson , A. (2000). Joint modeling of longitudinal measurements and event time data. Biostatistics, Vol. 1(4): 465-480.
13. . Guo, X. and Carlin, B.P. (2004). Separate andJoint modeling of longitudinal and event time data using standard computer packages. The American Statistician, Vol 58(1): 16-24.
14. Bowman, F.D. and Manatunga, A.K. (2005). A joint model for longitudinal data and associated event risks with application to a depression study. Applied statistics, Vol. 54(2): 301-316.
15. Stokes EM, Davis CS, Koch GG. *Categorical Data Analysis Using The SAS System (2000).*, SAS Inst. Inc.
16. Liang, K-Y and Zeger Sl. Longitudinal data analysis using generalized linear models. *Biometrika.* 1986, 73:13-22.
17. Zeger Sl and Liang K-Y. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics.* 1986, 42:121-130.
18. Statistics Canada (2002). Documentation of the Longitudinal Component of the National Population Health Survey.
19. Pahwa, P., Karunanayake and McDuffie H.H. (2006). Modeling of Longitudinal Polytomous Outcomes From Complex Survey Data. Joint Sttaistical Meeting Proceedings.

20. Kessler, R.C. and Morczek, D. (1994). Final version of our non-specific psychological distress scale (memo dated 10/3/94]. Ann Arbort (MI): Survey Research Centre of the Institute for Social Research. University of Michigan.
21. Wang, J. and EI-Guebaly N. (2004). Socio-demographic factors associated with co-morbid major depressive episodes and alcohol dependence in the general population. Canadian Journal of Psychiatry. Vol. 49(1):37-44.

Table:

	Model without drop out variable		Model with drop variable	
	$\hat{\beta}\,[s.e.(\hat{\beta})]$	OR (95% CI)	$\hat{\beta}\,[s.e.(\hat{\beta})]$	OR (95% CI)
Intercept	-3.28 [0.15]	0.04 (0.03,0.05)	-3.37 [0.15]	0.04 (0.03,0.05)
Drop[1]				
Died during six cycles			0.13 [0.05]	1.14 (1.02,1.26)
At least one missing				
Demographic Information				
Age Group[2]				
15-24 years	0.85 [0.12]	2.34 (1.85,2.97)	0.88 [0.12]	2.40 (1.89,3.05)
25-54 years	0.62 [0.08]	1.86 (1.60,2.17)	0.65 [0.08]	1.91 (1.63,2.23)
55-69 years	0.14 [0.08]	1.15 (0.98,1.34)	0.16 [0.08]	1.17 (1.00,1.37)
Ethnic groups[3]				
East Europeans	0.25 [0.30]	1.28 (0.70,2.32)	0.25 [0.30]	1.28 (0.71,2.33)
Western Europeans	0.23 [0.15]	1.26 (0.94,1.67)	0.23 [0.15]	1.26 (0.95,1.68)
Chinese	-0.73 [0.75]	0.48 (0.11,2.11)	-0.72 [0.75]	0.49 (0.11,2.10)
South Asian	-0.98 [0.62]	0.37 (0.11,1.25)	-0.99 [0.63]	0.37 (0.11,1.27)
Black	1.18 [0.61]	3.25 (0.98,10.82)	1.15 [0.60]	3.17 (0.98,10.30)
Other	0.18 [0.20]	1.20 (0.81,1.79)	0.17 [0.20]	1.19 (0.80,1.77)
Immigration Status[4]				
Immigrant	0.09 [0.09]	1.09 (0.92,1.30)	0.08 [0.09]	1.08 (0.91,1.29)
Social Role[5]				
Single parent, not employed	0.47 [0.13]	1.59 (1.23,2.06)	0.46 [0.13]	1.59 (1.22,2.06)
Single parent, employed	0.11 [0.10]	1.11 (0.91,1.36)	0.11 [0.10]	1.11 (0.92,1.36)
Partnered with children, not employed	0.01 [0.06]	1.01 (0.91, 1.13)	0.01 [0.06]	1.01 (0.91, 1.14)
Partnered with children, employed	-0.06 [0.06]	0.94 (0.83,1.07)	-0.05 [0.06]	0.95 (0.84,1.07)
Location of residence[6]				
Rural	-0.14 [0.05]	0.87 (0.78,0.96)	-0.14 [0.05]	0.87 (0.78,0.97)
Geographical area[7]				
Eastern	0.05 [0.08]	1.05 (0.90,1.23)	0.06 [0.08]	1.06 (0.90,1.24)
British Columbia	0.04 [0.08]	1.04 (0.88,1.22)	0.04 [0.08]	1.04 (0.88,1.23)
Central	-0.08 [0.07]	0.92 (0.80,1.06)	-0.08 [0.07]	0.93 (0.80,1.07)
Quebec	0.45 [0.07]	1.56 (1.36,1.80)	0.45 [0.07]	1.56 (1.36,1.80)
Length of stay in Canada[8]				
Less than 2 years	-0.14 [0.29]	0.87 (0.49,1.55)	-0.16 [0.30]	0.85 (0.48,1.52)
2 or more and less than 20 years	0.38 [0.09]	1.46 (1.22,1.76)	0.37 [0.09]	1.45 (1.21,1.75)
Socio-economic status				
Education level[9]				
Less or equal to 12 years	0.09 [0.09]	1.10 (0.92,1.31)	0.09 [0.0914]	1.09 (0.91,1.31)
Income level[10]				
Low	0.36 [0.08]	1.44 (1.24,1.67)	0.35 [0.08]	1.42 (1.22,1.66)
Middle	0.09 [0.06]	1.09 (0.97,1.23)	0.08 [0.06]	1.09 (0.97,1.22)
Social Support				
Social Involvement Score[11]				
Low	0.24 [0.11]	1.27 (1.02,1.58)	0.232 [0.11]	1.26 (1.01,1.56)
Moderate	0.10 [0.11]	1.11 (0.89,1.37)	0.10 [0.11]	1.10 (0.89,1.37)
Life-style				
Smoking Status[11]				
Current smoker	0.41 [0.12]	1.50 (1.18,1.91)	0.40 [0.12]	1.49 (1.17,1.89)
Ex-Smoker	0.27 [0.09]	1.31 (1.10,1.55)	0.27 [0.09]	1.31 (1.10,1.55)
Household Smoking[12]				
Yes	0.10 [0.10]	1.10 (0.92)	0.10 [0.10]	1.10 (0.91,1.33)
Health- Related:				
General Health status[13]				
Poor	2.71 [0.12]	15.00 (11.81,19.04)	2.69 [0.12]	14.78 (11.63,18.77)
Fair	1.80 [0.08]	6.05 (5.18,7.06)	1.79 [0.08]	6.01 (5.14,7.02)

Good	1.06 [0.06]	2.89 (2.55,3.28)	1.06 [0.06]	2.88 (2.54,3.27)
Very Good	0.42 [0.06]	1.53 (1.36,1.72)	0.42 [0.06]	1.53 (1.36,1.72)
Time point[14]				
Cycle 6	-0.30 [0.07]	0.74 (0.65,0.84)	-0.28 [0.07]	0.76 (0.66,0.87)
Cycle 5	-0.40 [0.07]	0.67 (0.59,0.76)	-0.38 [0.07]	0.68 (0.60,0.7876)
Cycle 4	-0.54 [0.06]	0.58 (0.52,0.66)	-0.52 [0.06]	0.59 (0.53,0.67)
Cycle 3	-0.31 [0.06]	0.74 (0.65,0.83)	-0.30 [0.06]	0.74 (0.66,0.83)
Cycle 2	-0.36 [0.05]	0.70 (0.63,0.77)	-0.36 [0.05]	0.70 (0.63,0.78)
Smoking * Ethnicity				
Current smoker *East Europeans	-0.76[0.37]	0.47 (0.23,0.97)	-0.76 [0.37]	0.47 (0.23,0.97)
Current smoker * Western Europeans	0.04 [0.16]	1.042 (0.76,1.43)	0.04 [0.16]	1.04 (0.76,1.43)
Current smoker * Chinese	0.21 [0.50]	1.23 (0.46,3.28)	0.22 [0.51]	1.24 (0.46,3.39)
Current smoker * South Asian	0.58 [1.20]	1.78 (0.17,18.69)	0.62 [1.21]	1.85 (0.17,20.02)
Current smoker * Black	-0.27 [0.42]	0.763 (0.34,1.74)	-0.30 [0.42]	0.74 (0.33,1.69)
Current smoker * Other	-0.14 [0.22]	0.87 (0.57,1.33)	-0.14 [0.22]	0.87 (0.57,1.34)
Ex-Smoker * East Europeans	-0.64 [0.28]	0.53 (0.30,0.92)	-0.65 [0.28]	0.52 (0.30,0.91)
Ex-Smoker * Western Europeans	-0.16 [0.12]	0.85 (0.67,1.09)	-0.16 [0.13]	0.85 (0.67,1.09)
Ex-Smoker * Chinese	0.20 [0.56]	1.23 (0.40,3.67)	0.21 [0.57]	1.23 (0.40,3.77)
Ex-Smoker * South Asian	-0.21 [0.62]	0.81(0.24,2.74)	-0.18 [0.64]	0.84 (0.24,2.91)
Ex-Smoker * Black	0.66 [0.62]	1.94 (0.57,6.58)	0.66 [0.62]	1.93 (0.57,6.49)
Ex-Smoker * Other	-0.10 [0.17]	0.90 (0.64,1.26)	-0.10 [0.17]	0.90 (0.64,1.27)
Non-Smoker * British	Reference		Reference	
Education*Ethnicity				
Less or equal to 12 years*East Europeans	0.27 [0.24]	1.31 (0.81,2.09)	0.26 [0.24]	1.30 (0.81,2.08)
Less or equal to 12 years*Western Europeans	-0.03 [0.12]	0.97 (0.77,1.22)	-0.04 [0.12]	0.96 (0.77,1.21)
Less or equal to 12 years *Chinese	0.47 [0.39]	1.60 (0.74,3.47)	0.49 [0.39]	1.63 (0.75,3.53)
Less or equal to 12 years * South Asian	-0.19 [0.54]	0.82 (0.29,2.37)	-0.21 [0.54]	0.81 (0.28,2.34)
Less or equal to 12 years *Black	-1.55 [0.57]	0.21 (0.07,0.65)	-1.55 [0.57]	0.21 (0.07,0.65)
Less or equal to 12 years * Other	0.08 [0.15]	1.08 (0.80,1.46)	0.08 [0.15]	1.08 (0.80,1.46)
Greater than 12 years *British	Reference		Reference	
Social Involvement Score * Ethnicity				
Low*East Europeans	-0.11 [0.34]	0.90 (0.46,1.74)	-0.09 [0.34]	0.91 (0.47,1.77)
Low* Western Europeans	-0.01 [0.16]	0.99 (0.72,1.35)	-0.01 [0.16]	0.99 (0.73,1.36)
Low* Chinese	0.02 [0.82]	1.02 (0.21,5.09)	0.01 [0.81]	1.01 (0.21,5.00)
Low* South Asian	0.99 [0.83]	2.69 (0.52,13.74)	0.99 [0.84]	2.68 (0.51,14.03)
Low* Black	-1.26 [0.64]	0.28 (0.08,1.00)	-1.22 [0.64]	0.30 (0.08,1.03)
Low* Other	-0.04 [0.22]	0.96 (0.62,1.49)	-0.04 [0.22]	0.97 (0.62,1.50)
Moderate* East Europeans	0.04 [0.29]	1.04 (0.58,1.85)	0.03 [0.29]	1.03 (0.58,1.84)
Moderate* Western Europeans	0.09 [0.16]	1.09 (0.80,1.48)	0.09 [0.16]	1.09 (0.80,1.48)
Moderate* Chinese	1.05 [0.77]	2.85 (0.62,12.98)	1.02 [0.77]	2.78 (0.62,12.51)
Moderate* South Asian	0.44 [0.74]	1.56 (0.36,6.70)	0.44 [0.75]	1.55 (0.36,6.79)
Moderate* Black	-2.65 [0.84]	0.07 (0.01,0.37)	-2.62 [0.82]	0.073 (0.01,0.37)
Moderate* Other	-0.03 [0.22]	0.97 (0.63,1.51)	-0.03 [0.22]	0.97 (0.62,1.50)
High* British	Reference		Reference	
Household Smoking*Ethnicity				
Household Smoking*East Europeans	0.76 [0.31]	2.13 (1.17,3.90)	0.76 [0.31]	2.14 (1.18,3.91)
Household Smoking*Western Europeans	0.07 [0.13]	1.07 (0.83,1.38)	0.07 [0.13]	1.08 (0.84,1.39)
Household Smoking *Chinese	0.07 [0.62]	1.07 (0.32, 3.59)	0.08 [0.62]	1.08 (0.32, 3.66)
Household Smoking * South Asian	-0.15 [1.08]	0.86 (0.10,7.17)	-0.14 [1.10]	0.87 (0.10,7.51)
Household Smoking *Black	0.74 [0.90]	2.09 (0.36,12.22)	0.76 [0.91]	2.14 (0.36, 12.64)
Household Smoking * Other	0.17 [0.17]	1.18 (0.84,1.66)	0.17 [0.17]	1.19 (0.85,1.67)
Household Smoking *British	Reference:		Reference	

Reference Categories: 1 Completers ; 2. 70 years and over; 3. British; 4 Non-immigrant; 5. Women without children, partnered and single; 6. Urban; 7. Ontario; 8. 20 or more years; 9 Greater than 12 years; 10. High; 11. High; 12.. Non-Smoker; 13. No; 14. Excellent; 15. Excellent; 16. Cycle 1