9-27-2017

Generalized Linear Mixed Model Analysis of Urban-Rural Differences in Social and Behavioral Factors for Colorectal Cancer Screening

Ke-Sheng Wang
East Tennessee State University

Xuefeng Liu

Muyiwa Ategbole

Xin Xie

Ying Liu

See next page for additional authors

Recommended Citation

Wang, K. S., Liu, X., Ategbole, M., Xie, X., Liu, Y., Xu, C., Xie, C., & Sha, Z. (2017). Generalized Linear Mixed Model Analysis of Urban-Rural Differences in Social and Behavioral Factors for Colorectal Cancer Screening. Asian Pacific journal of cancer prevention : APJCP, 18(9), 2581–2589. https://doi.org/10.22034/APJCP.2017.18.9.2581

RESEARCH ARTICLE

Generalized Linear Mixed Model Analysis of Urban-Rural Differences in Social and Behavioral Factors for Colorectal Cancer Screening

Ke-Sheng Wang[1]*, Xuefeng Liu[2], Muyiwa Ategbole[1], Xin Xie[3], Ying Liu[1], Chun Xu[4], Changchun Xie[5], Zhanxin Sha[6]

Abstract

Objective: Screening for colorectal cancer (CRC) can reduce disease incidence, morbidity, and mortality. However, few studies have investigated the urban-rural differences in social and behavioral factors influencing CRC screening. The objective of the study was to investigate the potential factors across urban-rural groups on the usage of CRC screening. **Methods:** A total of 38,505 adults (aged ≥40 years) were selected from the 2009 California Health Interview Survey (CHIS) data - the latest CHIS data on CRC screening. The weighted generalized linear mixed-model (WGLIMM) was used to deal with this hierarchical structure data. Weighted simple and multiple mixed logistic regression analyses in SAS ver. 9.4 were used to obtain the odds ratios (ORs) and their 95% confidence intervals (CIs). **Results:** The overall prevalence of CRC screening was 48.1% while the prevalence in four residence groups - urban, second city, suburban, and town/rural, were 45.8%, 46.9%, 53.7% and 50.1%, respectively. The results of WGLIMM analysis showed that there was residence effect ($p<0.0001$) and residence groups had significant interactions with gender, age group, education level, and employment status ($p<0.05$). Multiple logistic regression analysis revealed that age, race, marital status, education level, employment stats, binge drinking, and smoking status were associated with CRC screening ($p<0.05$). Stratified by residence regions, age and poverty level showed associations with CRC screening in all four residence groups. Education level was positively associated with CRC screening in second city and suburban. Infrequent binge drinking was associated with CRC screening in urban and suburban; while current smoking was a protective factor in urban and town/rural groups. **Conclusions:** Mixed models are useful to deal with the clustered survey data. Social factors and behavioral factors (binge drinking and smoking) were associated with CRC screening and the associations were affected by living areas such as urban and rural regions.

Keywords: Colorectal cancer- screening- mixed model- urban-rural differences- binge drinking- smoking

Asian Pac J Cancer Prev, **18 (9)**, 2581-2589

Introduction

Colorectal cancer (CRC) is the second leading cause of cancer death in the United States (US) (Jemal et al., 2010; Young and Womeldorph, 2013; CDC, 2014; Siegel et al., 2014; American Cancer Society, 2015). Screening for CRC can assist to identify and diagnose the disease at early stage; therefore it can reduce cancer incidence, morbidity, and mortality (Walsh and Terdiman, 2003; Espey et al., 2007; Xirasagar et al., 2015; Zauber, 2015). It has been suggested that sigmoidoscopy and colonoscopy have potential to both detect and treat CRC (Levin et al., 2008; Schoen et al., 2012); while colonoscopy has been considered as a primary screening test and a dominant CRC screening method in the US (Lieberman and Weiss, 2001; Young and Womeldorph, 2013).

Although CRC screenings are covered by the US Medicare program as one of the preventive services, the uptake of CRC screening is relatively low, and it is about 50% of those for whom the test is highly recommended (Seeff et al., 2004; Meissner et al., 2006; McGregor et al., 2007). Previous studies suggested several potential factors influencing CRC screening such as age, gender, educational level, income level, race, alcohol use, family history of CRC, health insurance status (Etzioni et al., 2004; Seeff et al., 2004; Wong et al., 2005; Beydiun and

*[1]Department of Biostatistics and Epidemiology, College of Public Health, [3]Department of Economics and Finance, College of Business and Technology, East Tennessee State University, Johnson City, TN 37614, [2]Department of Systems Leadership and Effectiveness Science, School of Nursing, University of Michigan, Ann Arbor, MI 48109-5482, [4]Department of Health and Biomedical Sciences, College of Health Affairs, University of Texas Rio Grande Valley, Brownsville, TX 78520, [5]Division of Biostatistics and Bioinformatics, Department of Environmental Health, University of Cincinnati, Cincinnati, OH 45267, [6]School of Kinesiology, College of Health, University of Southern Mississippi, Hattiesburg, MS 39406, USA. *For Correspondence: wangk@etsu.edu*

Beydoun, 2008; Johnson-Kozlow er al., 2009; Maxwell et al., 2010; López-Charneco et al., 2013; Modiri et al., 2013; Perencevich et al., 2013; Owusu et al., 2014, 2015), chronic disease conditions (hypertension, cancer history, arthritis, ulcer, and high cholesterol level) (Owusu et al., 2014), health behaviors and mental health (depression and insomnia) (Modiri et al., 2013; Owusu et al., 2015).

Correlated data are fairly common in health and social sciences. For example, clustered data arise when subjects are nested in clusters such as hospitals, regions, and neighborhoods; observations within the same cluster are likely to be correlated. Mixed models (also known as multilevel models or hierarchical models) including both fixed effects and random effects have been proposed to analyze correlated data (Li et al., 2011; West et al., 2014; Ene et al., 2015; Wang, 2016). The generalized linear mixed model (GLMM) is considered as an extension of the generalized linear model (e.g., logistic regression or Poisson regression) to include both fixed and random effects and have been used in analysis of complex survey data. For example, effect of metropolitan/micropolitan statistical areas was evaluated to estimate the relationship between occupational structure and the prevalence of coronary heart disease (Michimi et al., 2013). Other studies used the neighborhood as zip code when applied the GLMMs to investigate racial/ethnic disparities in oral health care (Brumback et al., 2013, 2014). However, a few studies have focused on the urban-rural differences in CRC screening (Benuzillo et al., 2009; Cole et al., 2012; Anderson et al., 2013; Davis et al., 2013; Modiri et al., 2013; Hughes et al., 2015); while no study has been found to use a mixed model analysis to deal with the hierarchical structure data and adjust for the sample weights in CRC screening.

This study aimed to determine the prevalence of CRC screening and the factors associated with CRC screening among adults aged ≥40 years living in California across urban and rural regions. We first used a weight generalized linear mixed model (WGLMM) to examine whether there is random effect among 4 dwelling regions. Then we used the WGLMM to detect associations of social and behavioral factors with CRC screening in the whole sample and to examine whether such associations differed by urban and rural regions.

Materials and Methods

Study population

The California Health Interview Survey (CHIS) is conducted by a collaborative study of the University of California, Los Angeles (UCLA) Center for Health Policy Research, the California Department of Health Services, and the Public Health Institute. The 2009 CHIS data was from the fifth CHIS data collection cycle since 2001. From each household, one adult respondent aged 18 years or older was randomly selected. Details about the sampling design can be found on the CHIS webpage (http://www.chis.ucla.edu/design.html). There was IRB exemption due to secondary data analysis.

Measurements

CRC screening. Subjects aged ≥ 40 years were considered to have had a CRC screening if they responded "yes" to either or both of the two questions "Have you ever had a colonoscopy?" and "Have you ever had a sigmoidoscopy?" (Table 1). Sigmoidoscopy and colonoscopy are medical examinations in which a tube is inserted in the rectum to view the colon for signs of cancer or other health problems. In total, 22,871 individuals with CRC screening of 38,505 adults (aged 40 or older) were available for the 2009 CHIS data, which is the latest CHIS data on CRC screening.

Social factors. Gender was coded as male or female based on self-report. Age was categorized as 40-49 years, 50-64 years, and 65 years or older. Race consisted of five subgroups: Whites, Latino, Asian, African American, and other. Employment status was dichotomized into either yes or no. Poverty level was categorized into four levels, including 0-99 % federal poverty level (FPL), 100-199 % FPL, 200-299 % FPL, and 300% FPL or above. Marital status was classified into married/living with partner, widowed/divorced/separated, and never married. Education indicated whether the participant had a high school's degree or not. Place of birth was coded as being born in the US or outside the US. Residence included four residence regions groups - urban, second city, suburban, and town/rural based on zip code.

Behavioral factors. Smoking status was categorized as never smoking, current smoking, and former smoking. Binge drinking was defined for women when the individuals had 4 or more drinks at one time during the last 30 days and for men when the individuals had 5 or more drinks at one time during the last 30 days. There are 3 categories: no binge drinking past year, infrequent binge drinking (less than monthly), and frequent binge drinking (monthly, daily or weekly).

Statistical analysis

Descriptive statistics and prevalence

The PROC CROSSTAB in SAS-Callable SUDAAN 11, which uses Taylor Series Linearization to account for the weighting of the data and the complex survey sampling strategy, was used to weight and estimate population proportions of CRC screening. The overall prevalence and prevalence for potential factors were estimated. The Chi-square test was used to compare the prevalence of CRC screening across age groups, gender, race, and other factors.

Weighted generalized linear mixed model (WGLIMM)

Let level-1 be the individual level and level-2 be the region level. Using a two-level hierarchical logistic regression model, the higher-level unit has its own intercept in the model, where the subject-specific intercepts are used to measure the differences among regions or neighborhoods. The two-level logistic regression model including random effects (such as region effect) for a binary dependent outcome can be extended to Equation (1), which is a logistic mixed model, a member of GLMMs.

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \alpha_j + \beta x_{ij}$$
$$\alpha_j = \alpha + u_j, u_j \sim N(0, \sigma^2) \quad (1)$$

where, i =1,…, I_j is level-1 individual i indicator, and j =1,…, J is the level-2 indicator such as region or neighborhood. The p_{ij} is the probability of outcome for individual *i* in the level *j*, conditional on the risk factor x. β is the vector of slopes. The region effects are measured by the random intercepts u_j, which are assumed to be normally distributed. α and β are fixed effects and u_j are random effects.

The PROC GLIMMIX in SAS version 9.4 was used to fit the logistic mixed model (e.g., Dai et al., 2006; Zhu, 2014; Ene et al., 2015; Wang, 2016). In the present study, the weighted generalized linear mixed-model (WGLIMM) regression analysis in PROC GLIMMIX was used to deal with the hierarchical structure data and also adjust for sampling weights. The RANDOM statement was used to specify intercept as random effects (u_j) of 4 regions in this data, while the WEIGHT statement was used to adjust for weights. In the WGLIMM analysis, the significance of the random effect (u_j) was detected using approximate z test while the fixed effects (α and β) were tested using an approximate *t* test. The ESTIMATE statement was used to test the difference in random effects between regions; while the odds ratio (OR) and its confidence interval (CI) for random effects were calculated. The univariate logistic mixed regression was used to examine the unadjusted association of each potential risk factor with CRC screening; while the multiple logistic mixed regression analysis was conducted to determine the adjusted associations of potential risk factors with CRC screening. The OR and its 95% CI for fixed effects were calculated using the ODDSRATIO option on the MODEL statement.

When considering interactions between region and other potential factors (Dai et al., 2006), the WGLIMM can be extended to Equation (2)

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \alpha_j + \gamma z_j + \beta x_{ij} + \theta z_j x_{ij}$$
$$\alpha_j = \alpha + u_j, u_j \sim N(0, \sigma^2) \quad (2)$$

where, z_j is an indicator for the 4 regions with fixed effect (γ). θ is the parameter for the interaction term $z_j x_{ij}$.

The WGLIMM analyses were performed using SAS statistical software, version 9.4 (SAS Institute, Cary, NC, USA).

Results

Prevalence of colorectal cancer screening

The prevalence of CRC screening is listed in Table 1. The overall prevalence of CRC screening was 48.1% (47.5% for males and 48.7 % for females). The prevalence increased with age (15.6%, 58.6% and 75.3% for age groups 40-49, 50-64 and 65 years, respectively). There is significant difference among the ethnic groups; the Whites revealed the highest prevalence (57%) compared with AA, Asian, and Hispanic (46.9%, 43.3%, 32%, respectively).

Table 1. Prevalence of Colorectal Cancer Screening (%)

Variable	Total (N)	Cases (N)	Prevalence (%)	*p value*
Gender				
Male	15,480	9,323	47.5 (45.9-48.9)	0.187
Female	23,025	13,548	48.7(47.6-49.8)	
Age group				
40-49	7,648	1,353	15.6(13.9-17.5)	<0.0001
50-64	15,292	9,427	58.6(56.9-60.3)	
65 +	15,565	12,091	75.3 (73.9-76.7)	
Race				
Hispanic	4,978	1,980	32.0 (30.0-34.0)	<0.0001
Whites	27,331	17,714	57.0 (55.9-58.1)	
AA	1,519	891	46.9 (42.3-51.6)	
Asian	3,576	1,694	43.3 (39.2-47.3)	
other	1101	592	43.3 (37.4-49.2)	
Marital status				
Married/living together	20,892	1,2439	49.9 (48.6-51.0)	<0.0001
Never married	3,387	1,591	33.3 (29.1-37.6)	
Other	14,226	8,841	48.3 (46.3-50.3)	
Education				
≤HS	11,227	5,856	40.7 (39.2-42.2)	<0.0001
>HS	27,078	16,931	53.2 (52.0-54.3)	
Born in US				
Yes	30,123	19,098	53.9 (52.9-55.0)	<0.0001
No	8,382	3,782	35.9 (33.8-38.2)	
Region				
Urban	11,635	6,591	45.8 (43.9-47.7)	<0.0001
2nd city	11,263	6,618	46.9 (45.3-48.5)	
Suburban	8,094	5,176	53.7 (51.6-55.7)	
Town/ rural	7513	4486	50.1 (48.1-52.0)	
Employment				
No	18,365	9,094	40.7 (39.3-42.1)	<0.0001
Yes	20,140	13,777	58.3 (56.7-59.8)	
Poverty level				
0-99% FPL	3,838	1,604	26.8 (23.6-30.3)	<0.0001
100-199% FPL	6,187	3,166	42.5 (39.7-45.4)	
200-299% FPL	5,205	3,057	46.9 (43.4-50.5)	
300% FPL +	23,275	15,044	54.9 (53.9-55.9)	
Binge drinking				
Never	31,156	19,204	51.2 (50.2-52.2)	<0.0001
Infrequent	4,294	2,168	39.2 (36.5-41.9)	
Frequent	3,055	1,499	35.2 (32.1-38.3)	
Smoking status				
Never	20,848	11,811	45.9 (44.6-47.3	<0.0001
Current	4,220	1,869	33.6 (31.1-36.0)	
Former	13,437	9,191	57.9 (56.4-59.5)	
Overall	38,505	22,871	48.1(47.2-49.1)	

p value is based on $\chi 2$ test; Abbreviations: AA, African American; HS, High school; FPL, federal poverty level

Table 2. Random Region Effects Based on a Two-Level Hierarchical Logistic Regression Model

Effect	Subject	Estimate	Standard error	*p value*
Intercept	Region	0.005356	0.00119	<0.0001
Intercept	Urban	0.02842	0.04116	0.4899
Intercept	2nd city	-0.07841	0.03842	0.0413
Intercept	Suburban	0.08769	0.03686	0.0174
Intercept	Town/rural	-0.08142	0.03642	0.0256

p value is based on approximate z test in the weighted generalized linear mixed model (WGLIMM) analysis

Table 3. Random Region Effects Comparison

Subject	OR	95%CI	*p value*
Suburban vs. Urban	1.18	1.17-1.20	<0.0001
Suburban vs. 2nd city	1.11	1.10-1.12	<0.0001
Suburban vs. Town/rural	1	0.99-1.01	0.157
Town/rural vs. Urban	1.18	1.17-1.19	<0.0001
Town/rural vs. 2nd city	1.12	1.11-1.13	<0.0001
2nd city vs. Urban	1.06	1.05-1.07	<0.0001

p value is based on approximate t test in the weighted generalized linear mixed model (WGLIMM) analysis

Table 4. Fixed Effects in Logistic Regression Analyses for the Relationship between Potential Factors and Colorectal Cancer Screening

Variable	Crude OR	95% CI	*p-value*	Adjusted OR	95% CI	*p value*
Gender						
Male	1			1		
Female	1.06	1.00-1.13	0.0699	0.93	0.85-1.02	0.1207
Age group						
40-49 years	1			1		
50-64 years	7.9	6.73-9.28	<0.0001	7.36	6.13-8.82	<0.0001
65 + years	17.4	15.4-19.8	<0.0001	15.04	12.69-17.84	<0.0001
Race						
Whites	1			1		
Hispanic	0.36	0.33-0.38	<0.0001	0.71	0.60-0.85	0.0002
AA	0.66	0.62-0.70	<0.0001	0.95	0.80-1.12	0.5177
Asian	0.54	0.51-0.58	<0.0001	0.69	0.58-0.83	<0.0001
Other	0.58	0.44-0.77	0.0001	0.73	0.52-1.03	0.0745
Marital status						
Never	1			1		
Married/living together	1.89	1.63-2.21	<0.0001	1.18	0.99-1.42	0.0655
Other	1.82	1.65-2.01	<0.0001	0.94	0.82-1.08	0.3849
Education						
>HS	1			1		
≤HS	0.62	0.57-0.67	<0.0001	0.74	0.65-0.85	<0.0001
Born in US						
No	1			1		
Yes	2.11	1.71-2.60	<0.0001	1.25	1.05-1.49	0.0106
Employment						
No	1			1		
Yes	2.08	1.95-2.22	<0.0001	1.4	1.28-1.52	<0.0001
Poverty level						
0-99% FPL	1			1		
100-199% FPL	1.99	1.72-2.31	<0.0001	1.53	1.39-1.67	<0.0001
200-299% FPL	2.43	2.18-2.70	<0.0001	2	1.94-2.06	<0.0001
300% FPL +	3.3	3.20-3.40	<0.0001	2.57	2.47-2.68	<0.0001
Binge drinking						
No	1			1		
Infrequent	0.61	0.49-0.76	<0.0001	0.91	0.79-1.04	0.1665
Frequent	0.51	0.44-0.59	<0.0001	0.76	0.67-0.87	<0.0001
Smoking status						
Never	1			1		
Current	0.6	0.54-0.67	<0.0001	0.63	0.48-0.82	0.0006
Former	1.62	1.18-1.65	<0.0001	1.07	1.00-1.14	0.0371

AA, African American; HS, High school; FPL, Federal Poverty Level; OR, odds ratio; CI, confidence interval

Table 5. Urban-Rural Differences in Multiple Logistic Regression Analyses for the Relationship between Potential Factors and Colorectal Cancer Screening

Variable	OR[1]	*p value*	OR[2]	*p value*	OR[3]	*p value*	OR[4]	*p value*
Gender								
Male	1		1		1		1	
Female	0.91 (0.75-1.10)	0.3126	0.79 (0.67-0.94)	0.0058	1.06 (0.88-1.29)	0.5275	1.09 (0.88-1.35)	0.4386
Age group								
40-49 years	1		1		1		1	
50-64 years	8.36 (6.58-10.6)	<0.0001	12.14 (9.2-16.0)	<0.0001	9.11 (7.18-11.6)	<0.0001	4.84 (3.67-6.37)	<0.0001
65 + years	16.8 (12.6-22.4)	<0.0001	6.2 (5.01-7.68)	<0.0001	19.5 (14.6-25.9)	<0.0001	11.6 (8.44-15.9)	<0.0001
Race								
Whites	1		1		1		1	
Hispanic	0.60 (0.45-0.80)	0.0004	0.95 (0.62-1.47)	0.8279	0.67 (0.47-0.96)	0.0282	0.76 (0.56-1.05)	0.092
AA	0.82 (0.56-1.20)	0.3086	0.84 (0.65-1.08)	0.1706	1.02 (0.62-1.68)	0.9318	1.06 (0.67-1.07)	0.8104
Asian	0.58 (0.42-0.63)	0.0011	0.84 (0.57-1.26)	0.4073	0.84 (0.60-1.18)	0.3093	1.05 (0.28-3.99)	0.9457
Other	0.48 (0.37-0.63)	0.0093	1.07 (0.66-1.74)	0.7869	0.72 (0.42-1.23)	0.224	1.34 (0.67-2.72)	0.4103
Marital status								
Never	1		1		1		1	
Married/ living together	1.02 (0.69-1.51)	0.9149	1.17 (0.76-1.81)	0.4796	1.64 (1.06-2.55)	0.0266	2.03 (1.33-3.10)	0.001
Other	0.84 (0.55-1.27)	0.411	1.01 (0.64-1.60)	0.9574	1.30 (0.80-2.10)	0.2845	1.21 (0.79-1.85)	0.378
Education								
>HS	1		1		1		1	
≤HS	0.82 (0.65-1.04)	0.1023	0.70 (0.58-0.84)	0.0002	0.58 (0.45-0.73)	<0.0001	0.85 (0.70-1.04)	0.117
Born in US								
No	1		1		1		1	
Yes	1.14 (0.88-1.47)	0.3333	1.72 (1.33-2.25)	<0.0001	1.20 (0.91-1.58)	0.2023	1.10 (0.79-1.53)	0.5558
Employment								
No	1		1		1		1	
Yes	1.38 (1.10-1.73)	0.0061	1.49 (1.23-1.80)	<0.0001	1.17 (0.96-1.44)	0.1277	1.75 (1.41-2.17)	<0.0001
Poverty								
0-99% FPL	1		1		1		1	
100-199% FPL	1.64 (1.13-2.38)	0.0092	1.30 (0.89-1.90)	0.168	1.71 (1.03-2.82)	0.0382	1.45 (0.94-2.26)	0.0966
200-299% FPL	1.97 (1.30-2.98)	0.0014	1.90 (1.25-2.89)	0.0026	2.35 (1.43-3.84)	0.0007	1.93 (1.23-3.01)	0.004
300% FPL+	2.49 (1.73-3.59)	<0.0001	2.54 (1.79-3.62)	<0.0001	2.95 (1.84-4.72)	<0.0001	2.67 (1.74-4.10)	<0.0001
Binge drinkng								
No	1		1		1		1	
Infrequent	0.78 (0.61-0.99)	0.0391	0.86 (0.64-1.17)	0.3442	0.59 (0.42-0.84)	0.0029	0.85 (0.64-1.13)	0.2615
Frequent	0.75 (0.56-1.02)	0.0646	1.03 (0.81-1.31)	0.8212	0.95 (0.72-1.26)	0.7232	0.98 (0.73-1.31)	0.8789
Smoking status								
Never	1		1		1		1	
Current	0.48 (0.35-0.67)	<0.0001	0.75 (0.56-1.02)	0.0652	0.86 (0.61-1.21)	0.3808	0.70 (0.53-0.93)	0.0151
Former	1.01 (0.84-1.22)	0.9216	1.10 (0.92-1.31)	0.3095	1.12 (0.91-1.38)	0.2801	1.19 (0.96-1.47)	0.1051

AA, African American; HS, High school; FPL, Federal Poverty Level; OR, odds ratio; CI, confidence interval; [1], refers to urban; [2], refers to second city; [3], refers to suburban; [4], refers to Town/rural.

The prevalence of CRC screening was highest in former smokers (57.9%) compared to never smokers (45.9%) and current smokers (33.6%). Never binge drinker recorded the highest prevalence of CRC screening compared to infrequent and frequent drinkers (51.2%, 39.2% and 35.2%, respectively). In the 4 residence regions, suburban residents had the highest prevalence comparing with those of urban, second city, and town/rural (53.7%, 45.8%, 46.9% and 50.1%, respectively).

Random region effects

Based on the WGLIMM analysis, the estimated covariance parameters and random region effects ($u_j^{\wedge}$) are shown in Table 2. The estimated variance of the

intercept was 0.005534 with a standard error of 0.001313, which measures the variability among 4 regions and the estimated value was significantly larger than 0 ($p<0.0001$) indicating that there was region effect on CRC screening. Furthermore, the random effects for second city, suburban, and town/rural were significant ($p=0.0413$, 0.0174, and 0.0256, respectively). Table 3 presents the comparison of random effects between regions. Residents in suburban and Town/rural revealed increased odds of CRC screening comparing with urban and second city regions; while people in second city had increased odds of screening comparing with urban. However, there was no significant difference in CRC screening between people living in suburban and town/rural.

Logistic mixed model for the whole sample

Table 4 presents the fixed effect results ($\hat{\beta}$) from univariate and multiple logistic mixed regression analyses considering random effects and weights. By using univariate analyses, all potential factors except for gender were associated with CRC screening ($p<0.05$). After adjusting for other factors using multiple logistic regression, age groups (50-64, and 65+), being born in the US, being employed, and former smoking were positively significantly associated with CRC screening (OR=7.36, 95%CI=6.13-8.82; OR=15.04, 95%CI=12.69-17.84; OR=1.15, 95%CI=1.05-1.49; OR=1.40, 95%CI=1.28-1.52; respectively); while lower education, frequent binge drinking, and current smoking was negatively associated with CRC screening (OR=0.74, 95%CI=0.65-0.85; OR=0.76, 95%CI=0.67-0.87; OR=0.63, 95%CI=0.48-0.82, respectively). In addition, high poverty levels were positively associated with CRC screening (OR=1.53, 95%CI=1.39-1.67; OR=2.00, 95%CI=1.94-2.06; OR=2.57, 95%CI=2.47-2.68 for 100-199% FPL, 200-299 % FPL, and 300% FPL or above, respectively).

Interactions with regions

After adjusting for potential risk factors in the multiple WGLIMM analysis, region showed significant interactions ($\hat{\theta}$) with gender ($p=0.0289$), age group ($p=0.0074$), education ($p=0.0444$), being born in the US ($p=0.0237$), and employment ($p=0.0239$), respectively.

Urban-rural differences

The urban-rural differences in the associations of potentials factors with CRC screening are shown in Table 5. Older age groups and higher poverty levels showed positively associations with CRC screening in all 4 residence groups, while being employed was a factor in all regions but not in suburban. Gender and being born in the US were factors just in second city (OR=0.79, 95%CI=0.67-0.94; OR=1.72, 95%CI=1.33-2.25, respectively). Hispanic people were negatively associated with CRC screening relative to whites in urban and suburban regions (OR=0.60, 95%CI=0.45-0.80; OR=0.67, 95%CI=0.47-0.96, respectively); while Asian people were less likely to use CRC screening than whites just in urban area (OR=0.58, 95%CI=0.42-0.63). Married people showed higher probability of using CRC screening in suburban and town/rural residence areas (OR=1.81, 95%CI=1.15-2.84; OR=2.57, 95%CI=1.67-3.92, respectively). Education level was positively associated with 2nd city and suburban (OR=0.70, 95%CI=0.58-0.84; OR=0.58, 95%CI=0.45-0.73, respectively).

Discussion

Using a large population-based study of CRC screening, we found that the overall prevalence of CRC screening in California adults was 48.1%; while there were significant differences among four residence groups - urban, second city, suburban, and town/rural. After adjusting for potential factors, age, race, marital status, education level, employment stats, binge drinking and smoking status were associated with CRC screening. Stratified by residence, age and poverty showed associations with CRC screening in all 4 residence groups. Gender and being born in US were factors just in 2nd city. Married people showed higher screening in suburban and town/rural residences. Education level was positively associated with 2nd city and suburban. Employment was a factor in all regions except for suburban region. Infrequent binge drinking was associated with CRC screening in urban and suburban; while current smoking was a potential factor in urban and town/rural groups.

In accordance with previous reports finding urban-rural differences in the prevalence of CRC screening ((Benuzillo et al., 2009; Cole et al., 2012; Anderson et al., 2013; Davis et al., 2013; Modiri et al., 2013; Hughes et al., 2015); our results further identified that the difference in prevalence among 4 types of residence regions was statistically significant. Using the CHIS data from 2001-2009, Modiri et al., (2013) reported the proportions of CRC screening using colonoscopy for people (aged 50-80 years) in these 4 regions (37.6%, 19.4%, 28.2% and 14.8% for urban, second city, suburban, and town/rural, respectively); however, they did not estimate prevalence of CRC screening in these regions. Using the CHIS 2009 data, we found that the proportions of CRC screening for people (aged ≥40 years) in these 4 regions are 41.7%, 24.2%, 22.3% and 11.8% for urban, second city, suburban, and town/rural, respectively (data not shown).

Previous studies have shown that CRC screening rate is lower in women than men (Green 1999; Etzioni et al., 2004; Seeff et al., 2004; Wong et al., 2005, Beydoun et al., 2008, Modiri et al., 2013; Owusu et al., 2014, 2015). However, the present study found that the prevalence of CRC screening in males and females in the whole sample did not show significant differences. Adjusting for all potential variables, gender was not associated with CRC screening in the whole sample. However, stratified analysis revealed that CRC screening was significantly higher in men than women in 2nd city. Furthermore, our results showed that CRC screening uptake increased with age as previous studies reported (Seeff et al., 2004; Beydiun and Beydoun, 2008; López-Charneco et al., 2013). Screening uptake was highest in those above 64 years old. The age difference may be explained by the fact that CRC risk increases with age and screening is recommended for those aged 50 years and above in the

US (Winawer et al., 2003). Our findings further added that in the 2nd city, the age group 50-64 years revealed the highest screening rate.

In terms of race, the present results showed CRC screening uptake was lower in all race groups compared to Whites; while all other races were significantly different from whites in CRC screening uptake using the univariate logistic regression. This finding was consistent with earlier studies (Seeff et al., 2004; Johnson-Kozlow er al., 2009; Maxwell et al., 2010; Perencevich et al., 2013; Johnson-Jennings et al., 2014) but contrast to our recent results (Owusu et al., 2014). In addition, we added new findings that the uptake rate in AA is not significantly different in the whole sample and in each subsample stratified by region; whereas Hispanic population was significantly lower just in Urban and suburban regions and Asian was also significantly lower in the urban region only.

Consistent with previous studies, being married or living with a partner increases the chance of being screened for CRC (Etzioni et al., 2004; Wong et al., 2005; Beydiun and Beydoun, 2008). However, our results further indicated that after adjusting for other factors, there is no significant difference in terms of marital status in CRC screening in the whole sample; whereas being married or living with a partner may increase the chance of being screened for CRC only in suburban and town/rural regions. Such difference in screening uptake may be potentially explained by the fact that couples gain support from their partners.

Our findings supported previous results that higher education, being employed and higher poverty level were associated with higher CRC screening (Lantz et al., 1997; Etzioni et al., 2004; Seeff et al., 2004; Wong et al., 2005; Beydiun and Beydoun, 2008; Johnson-Kozlow er al., 2009; Maxwell et al., 2010; López-Charneco et al., 2013; Owusu et al., 2014, 2015). It has been reported that education is well known to be a significant determinant of health; while education level of an individual may influence the level of understanding of CRC and the benefits of screening (Seeff et al., 2004; Beydoun et al., 2008; López-Charneco et al., 2013). Such insight is more likely to drive a person to accept and undergo screening (Lantz et al., 1997). Furthermore, our results further added that higher education was positively associated with CRC screening just in 2nd city and town/rural regions but not in urban and suburban regions; while employment was not associated with CRC screening in suburban areas. In addition, we found that being born in the US was associated with higher intake of CRC screening; whereas such association existed just in 2nd city region.

Previous studies have shown that moderate alcohol drinking was associated with increased cancer screening (Fredman et al., 1999; Owusu et al., 2015). However, our results showed that frequently binge drinking was negatively associated with CRC screening; whereas infrequently drinking was not associated with CRC screening after adjusting for other factors. Furthermore, our results showed that infrequently binge drinking was negatively associated with CRC screening in urban and suburban regions; whereas frequently drinking was not associated with CRC screening in any region. The differences between our results and previous reports may be due to the different alcohol scales used.

In accordance with previous reports, we found that former smoking was associated with increasing rate of CRC screening in the whole sample (Meissner et al., 2006; Coups et al., 2007; Owusu et al., 2015); whereas current smokers were less likely to participate in cancer screenings (Hama et al., 2016; Owusu et al., 2015). Stratification results indicated that former smoking was not associated with CRC screening in any of the 4 regions; while current smoking had lower odds for CRC screening just in urban and town/rural regions.

In our study, we used a large sample with diversity of the populations by use of five languages (English, Spanish, Chinese (Mandarin and Cantonese dialects), Vietnamese, and Korean) to cover the largest number for those who were able to neither speak English nor speak English well enough to otherwise participate. Furthermore, a large sample data of subjects were widely selected at random with comprehensive information for the wide age range on CRC screening and behavioral/health characteristics, which allowed us to adjust for numerous factors and give us a relatively large statistical power in estimations. In addition, the weighted generalized linear mixed-model (WGLIMM) regression analysis was used to deal with the hierarchical structure data (random region effects) and to adjust for weights. To the best of our knowledge, this is the first study to investigate urban-rural differences using a mixed model. However, the cross-sectional design could not determine a temporal or causal relationship between potential factors and CRC screening. Furthermore, self-reported data are subjective to some degree and may lead to misclassification. Moreover, the state-based data may limit the generalization of our findings. In addition, the analysis is limited by only using one year's data.

In conclusion, the GLIMMIX procedure in SAS can be used to perform weighted generalized linear mixed-model regression analysis dealing with the hierarchical structure data (random region effects) and adjusting for sampling weights. Social factors and behavioral factors (alcohol consumption and smoking) were associated with CRC screening; whereas the associations were affected by the living areas such as urban and rural regions. To improve CRC screening uptake, it is essential to consider urban-rural differences in predictors and tailor appropriate interventions to each region.

Role of the funding sources

No founding source is given for the present paper.

Conflict of interest

The authors declare that they have no conflict of interest.

Acknowledgements

The authors are grateful to the support of Data from the 2009 California Health Interview Survey.

References

American Cancer Society (2016). Key statistics for colorectal cancer. Atlanta (GA): American Cancer Society. http://www.cancer.org/acs/groups/cid/documents/webcontent/003096-pdf.

Anderson AE, Henry KA, Samadder NJ, Merrill RM, Kinney AY (2013). Rural vs urban residence affects risk-appropriate colorectal cancer screening. *Clin Gastroenterol Hepatol*, **11**, 526-33

Benuzillo JG, Jacobs ET, Hoffman RM, et al (2009). Rural-urban differences in colorectal cancer screening capacity in Arizona. *J Community Health*, **34**, 523-8.

Beydoun HA, Beydoun MA (2008). Predictors of colorectal cancer screening behaviors among average-risk older adults in the United States. *Cancer Causes Control*, **19**, 339-59.

Brumback BA, Cai Z, Dailey AB (2014). Methods of estimating or accounting for neighborhood associations with health using complex survey data. *Am J Epidemiol*, **179**, 1255-63.

Brumback BA, Zheng HW, Dailey AB (2013). Adjusting for confounding by neighborhood using generalized linear mixed models and complex survey data. *Stat Med*, **32**, 1313-24

Center for Disease Control and Prevention. Colorectal (colon) Cancer (2014). Available at: http://www.cdc.gov/cancer/colorectal/statistics/index.htm. [Accessed on 07/11/2014].

Cole AM, Jackson JE, Doescher M (2012). Urban-rural disparities in colorectal cancer screening: cross-sectional analysis of 1998-2005 data from the centers for disease control's behavioral risk factor surveillance study. *Cancer Med*, **1**, 350-6.

Coups EJ, Manne SL, Meropol NJ, Weinberg DS (2007). Multiple behavioral risk factors for colorectal cancer and colorectal cancer screening status. *Cancer Epidemiol Biomarkers Prev*, **16**, 510-6.

Dai J, Li Z, Rocke D (2006). Hierarchical logistic regression modeling with SAS GLIMMIX. SAS conference proceedings: Western users of SAS Software: 2006; Irvine, CA. 2006, http://www.lexjansen.com/cgi-bin/xsl_transform.php?x=wuss2006&c=wuss.

Davis TC, Rademaker A, Bailey SC, et al (2013). Contrasts in rural and urban barriers to colorectal cancer screening. *Am J Health Behav*, **37**, 289-98.

Ene M, Leighton EA, Blue GL, Bell BA (2015). Multilevel models for categorical data using SAS ® PROC GLIMMIX: The basics. SAS global forum 2015, Dallas, Texas. Available at http://support.sas.com/resources/papers/proceedings15/3430-2015.pdf.

Espey DK, Wu XC, Swan J, et al (2007). Annual report to the nation on the status of cancer, 1975-2004, featuring cancer in American Indians and Alaska Natives. *Cancer*, **110**, 2119-52.

Etzioni DA, Ponce NA, Babey SH, et al (2004). A population-based study of colorectal cancer test use: results from the 2001 California health interview survey. *Cancer*, **101**, 2523-32.

Fredman L, Sexton M, Cui Y, et al (1999). Cigarette smoking, alcohol consumption, and screening mammography among women ages 50 and older. *Prev Med*, **28**, 407-17.

Green CA, Pope CR (1999). Gender, psychosocial factors and the use of medical services: a longitudinal analysis. *Soc Sci Med*, **48**, 1363-72.

Hama H, Tabuchi T, Ito Y, et al (2006). Smoking behavior and participation in screening for lung, gastric, and colorectal cancers. *Nihon Koshu Eisei Zasshi*, **63**, 126-34.

Hughes AG, Watanabe-Galloway S, Schnell P, Soliman AS (2015). Rural-urban differences in colorectal cancer screening barriers in Nebraska. *J Community Health*, **40**, 1065-74

Jemal A, Siegel R, Xu J, Ward E (2010). Cancer statistics, 2010. *CA Cancer J Clin*, **60**, 277–300.

Johnson-Jennings MD, Tarraf W, Xavier Hill K, González HM (2014). United States colorectal cancer screening practices among American Indians/Alaska Natives, blacks, and non-Hispanic whites in the new millennium (2001 to 2010). *Cancer*, **120**, 3192-9.

Johnson-Kozlow M, Roussos S, Rovniak L, Hovell M (2009). Colorectal cancer test use among Californians of Mexican origin: influence of language barriers. *Ethn Dis*, **19**, 315-22.

Lantz PM, Weigers ME, House JS (1997). Education and income differentials in breast and cervical cancer screening: Policy implications for rural women. *Med Care*, **35**, 219-36.

Levin B, Lieberman DA, Mcfarland B, et al (2008). Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps, 2008: A joint guideline from the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology. *Gastroenterology*, **134**, 1570–95.

Li B, Lingsma HF, Steyerberg EW, Lesaffre E (2011). Logistic random effects regression models: a comparison of statistical packages for binary and ordinal outcomes. *BMC Med Res Methodol*, **11**, 77.

Lieberman DA, Weiss DG (2001). Veterans affairs cooperative study groupt one-time screening for colorectal cancer with combined fecal occult-blood testing and examination of the distal colon. *N Engl J Med*, **345**, 555–60.

López-Charneco M, Pérez CM, Soto-Salgado M, et al (2013). Correlates of colorectal cancer screening among Hispanics: Results from the 2008 Puerto Rico behavioral risk factor surveillance system survey. *P R Health Sci J*, **2**, 68-75.

Maxwell AE, Crespi CM, Antonio CM, Lu P (2010). Explaining disparities in colorectal cancer screening among five Asian ethnic groups: a population-based study in California. *BMC Cancer*, **10**, 214.

McGregor SE, Hilsden RJ, Li FX, Bryant HE, Murray A (2007). Low uptake of colorectal cancer screening 3 yr after release of national recommendations for screening. *Am J Gastroenterol*, **102**, 1727-35.

Meissner HI, Breen N, Klabunde CN, Vernon SW (2006). Patterns of colorectal cancer screening uptake among men and women in the United States. *Cancer Epidemiol Biomarkers Prev*, **15**, 389-94.

Michimi A, Ellis-Griffith G, Nagy C, Peterson T (2013). Coronary heart disease prevalence and occupational structure in U.S. metropolitan areas: a multilevel analysis. *Health Place*, **21**, 192-204.

Modiri A, Makipour K, Gomez J, Friedenberg F (2013). Predictors of colorectal cancer testing using the California Health Inventory Survey. *World J Gastroenterol*, **19**, 1247-55.

Owusu D, Longcoy J, Quinn M, Wang KS (2014). The relationship between chronic disease conditions and colon cancer screening: Results from the 2012 National Health Interview Survey. *Am J Cancer Epidemiol Prev*, **2**, 32-42.

Owusu D, Quinn M, Wang KS (2015). Alcohol consumption, depression, insomnia and colon cancer screening: Racial differences. *Int J High Risk Behav Addict*, **4**, e23424.

Perencevich M, Ojha RP, Steyerberg EW, Syngal S (2013). Racial and ethnic variations in the effects of family history of colorectal cancer on screening compliance. *Gastroenterology* **145**, 775-81.

Schoen RE, Pisky PF, Weissfeld JL, et al (2012). Colorectal-cancer incidence and mortality with screening flexible sigmoidoscopy. *N Engl J Med*, **366**, 2345–57.

Seeff LC, Nadel M R, Klabunde CN, et al (2004). Patterns

and predictors of colorectal cancer test use in the adult US population. *Cancer*, **100**, 2093-103.
Siegel R, Desantis C, Jemal A (2014). Colorectal cancer statistics, 2014. *CA Cancer J Clin*, **64**, 104–17.
Walsh JM, Terdiman JP (2003). Colorectal cancer screening: scientific review. *JAMA*, **289**, 1288-96.
Wang KS (2016). Linear and non-linear mixed models in longitudinal studies and complex survey data. *J Biom Biostat*, **7**, e290.
West BT, Welch KB, Galecki AT (2014). Linear mixed models: A practical guide using statistical software, Second Edition 2nd Edition. Publisher: Chapman and Hall/CRC; 2 edition. ISBN-10:1466560991, ISBN-13:978-1466560994.
Winawer S, Fletcher R, Rex D, et al (2003). Colorectal cancer screening and surveillance: clinical guidelines and rationale—update based on new evidence. *Gastroenterology* **124**, 544-60
Wong ST, Gildengorin G, Nguyen T, Mock J (2005). Disparities in colorectal cancer screening rates among Asian Americans and non-Latino whites. *Cancer*, **104**, 2940-7.
Xirasagar S, Li YJ, Hurley TG, et al (2015). Colorectal cancer prevention by an optimized colonoscopy protocol in routine practice. *Int J Cancer*, **136**, 731–42.
Young PE, Womeldorph CM (2013). Colonoscopy for colorectal cancer screening. *J Cancer*, **4**, 217-26.
Zauber AG (2015). The impact of screening on colorectal cancer mortality and incidence: has it really made a difference?. *Dig Dis Sci*, **60**, 681-91.
Zhu M (2014). Analyzing multilevel models with the GLIMMIX procedure. Cary, NC: SAS Institute Inc; 2014. Paper SAS026–2014. Read More: http://ajph.aphapublications.org/doi/full/10.2105/AJPH.2015.302842.