Estimation of Provincial Tobacco Use in Thailand Using Multilevel Small Area Estimation from a National Probability Sample

Jason Hsia[1,*], Xingyou Zhang[1], Hataichanok Puckcharern[2]
[1] US Centers for Disease Control and Prevention, Atlanta, USA
[2] National Statistical Office, Bangkok, Thailand
Corresponding author: Jason Hsia, e-mail: JHsia@cdc.gov

Abstract

Monitoring tobacco use is one of the key tobacco control strategies proposed by World Health Organization to assess implementation of the Framework Convention on Tobacco Control. In Thailand, tobacco use surveillance is based on repeated nationally representative cross sectional surveys. While these surveys provide national estimates of tobacco use, they do not provide provincial level estimates because they were not designed with adequate provincial sample sizes. Provincial estimates of tobacco use are necessary for assessing local tobacco control policies. We used small area estimation (SAE) to estimate prevalence of current tobacco use in 76 provinces and Bangkok in Thailand with information from a national probability sample and census data. Multilevel logistic regression models with a provincial-level random effect were first estimated based on the Global Adult Tobacco Survey in Thailand and then applied to 2010 Thailand census data, i.e., aggregated number of people with demographic characteristics, to make predictions of current tobacco use for all target population groups in each province. The multilevel model provincial-level estimates were evaluated and validated by comparison to their direct survey estimates. Variation of prevalence of current tobacco use at the provincial level was large among province. The correlation between estimated prevalence using SAE and direct estimation was 0.966 and 0.919 for men and women respectively. In conclusion: Application of SAE to a Thailand nationally representative sample can be used for reliable estimation of prevalence of current tobacco use for all 76 provinces and Bangkok in Thailand.

Key words: Small area estimation, random effect model, prevalence, tobacco use, Thailand

1. Introduction

Tobacco use is the leading preventable cause of death. If current use pattern continues, by 2030 there will be more than 8 million deaths annually attributable to tobacco smoke, an increase of 2 million from the current estimate (WHO, accessed on February 11, 2015). In responding to the public health crises, the World Health Organization's (WHO) Framework Convention on Tobacco Control (FCTC) came into force in February 2005, and it now has 176 countries who have committed to protect the health of their populace by implementing a set of tobacco control measures including monitoring tobacco use.

In Thailand, tobacco use surveillance is based on repeated nationally representative cross-sectional surveys (Hsia and Puckcharern, 2013). While these surveys provide national estimates of tobacco use prevalence, they do not provide provincial level estimates because they were not designed with adequate sample sizes to estimate target population parameters at the provincial level. In tobacco control practice, policy and tobacco control campaign often come from the provincial level. Therefore, provincial estimates of tobacco use are necessary for assessing local tobacco control policies. We used small area estimation (SAE) method to estimate prevalence of current tobacco use in 76 provinces and Bangkok metropolitan area in Thailand with information from a national probability sample and census data. We also evaluated this SAE approach used in the national representative tobacco surveys.

2. Methods

We applied the two-step method (Zhang et al, 2014) to obtain the provincial (small area) estimation for men and women current tobacco use (defined as all forms of tobacco use), respectively, as the prevalence rate among men

is much higher than that among women in Thailand. In the first step, we estimated parameters from following multilevel logistic regression models with a provincial-level random effect. The data used was from the Global Adult Tobacco Survey in Thailand (WHO-SEARO, 2011). $Logit\{P_{a,r,i,u,p}(y=1)\} = \alpha_a x_1 + \beta_r x_2 + \gamma_i x_3 + \tau_u x_4 + \mu_p z + \epsilon_{a,r,i,u,p}$, where y was self-reported status of current tobacco use, $\alpha_a, \beta_r, \gamma_i$, and τ_u represented parameters for fixed effects at individual level, age, region, income, and urbanicity, respectively, μ_p represented the parameter for random effect at provincial level, and $\epsilon_{a,r,i,u,p}$ was the error term. The random effect was assumed to be normally distributed. In the second step, we used the predicted value $\hat{P}_{a,r,i,u,p}(y=1) = Logit^{-1}\{\hat{\alpha}_a x_1 + \hat{\beta}_r x_2 + \hat{\gamma}_i x_3 + \hat{\tau}_u x_4 + \hat{\mu}_p z\}$ and 2010 Thailand census data, i.e., aggregated number of people with demographic characteristics, to make predictions of current tobacco use for all target population groups in each province. First, obtained the predicted number of current tobacco users in each province (N_{dp}), by summing products of predicted probabilities and the total number of men or women ($N_{a,r,i,u,p}$) in each of $a \times r \times i \times u$ categories for each province, i.e., $N_{dp} = \sum_{a,r,i,u} \hat{P}_{a,r,i,u,p}(y=1) \times N_{a,r,i,u,p}$. Second, we used N_{dp}/N_p to be the provincial estimator, where N_p was the total number of men or women in a province. In order to evaluate the obtained provincial-level estimates from the multilevel model approach, the estimates were compared to their direct estimates, taking complex survey feature into account.

3. Results

There was a large variation of prevalence of current tobacco use at the provincial level among provinces for both men and women, as shown in the Figure. For men, the range was from 24.4% in Lamphun to 68.2% in Ranong (median= 46.5%). For women, the range was from 1.41% in Narathiwat to 24.7% in Buri Ram (median=6.46%). The standard deviation of the prevalence ranged from 0.02% to 0.06% (median=0.04%) for men and from 0.01% to 0.05% (median=0.03%) for women. Table 1 shows the distribution of the estimates in details.

Table 2 provides the correlation between model-based estimated prevalence using SAE and direct survey estimates. The correlation was 0.966 and 0.919 for men and women respectively for weighted data.

4. Discussion

Thailand 2011 Global Adult Tobacco Survey followed standard global protocol and used a typical multistage stratified cluster sampling. Application of SAE method (Zhang et al. 2014) to such a nationally representative sample survey produced reliable estimates of current tobacco use prevalence for all 76 provinces and Bangkok metropolitan area in Thailand. This result was encouraging as it might be able to be applied to similar national surveys conducted in the other countries to get provincial/state level estimates. Global Adult Tobacco Survey protocol requires a sample size of 4,000 to estimate the smoking prevalence for desired target population (Global Adult Tobacco Survey Collaborative Group, 2010). Monitoring tobacco use needs repeated survey estimates at lower administrative level than national. Application of the SAE approach to get reliable estimates is a cost-effective solution to this problem. We also noticed that unlike some of existing SAE methods, the method we used was conceptually intuitive and was feasible to be implemented. This two-step SAE method only requires multilevel random effect modeling and poststratification-alike prediction.

The SAE approach we used also had its limitation. SAE results heavily relied on multilevel models used in the first step. With different covariates included in the model, estimates of the prevalence were different. This might be a challenge to apply it in tobacco use surveillance studies that is over time. Future evaluation and research are needed to the application of this SAE approach to the surveillance studies.

References

World Health Organization: Why tobacco is a public health priority http://www.who.int/tobacco/health_priority/en/

Hsia J. Puckcharern H. Comparisons between Cigarette Smoking and Alcoholic Drinking Behavior Survey and Global Adult Tobacco Survey in Thailand. Proceedings 59th ISI World Statistics Congress, 25-30 August 2013, Hong Kong (Session CPS204), P5685.

Zhang, X, Holt JB, Lu H, Weaton AG, Ford ES, Greenlund KJ, Croft JB. Multilevel regression and poststratification for small-area estimation of population health outcomes: A case study of chronic obstructive pulmonary disease prevalence using the Behavioral Risk Factor Surveillance System. Am. J. Epidemiol. (2014) 179 (8): 1025-33.

World Health Organization, Regional Office for South-East Asia (WHO-SEARO). Global Adult Tobacco Survey: Thailand report 2011. WHO Publication (2011).

Global Adult Tobacco Survey Collaborative Group. Global Adult Tobacco Survey (GATS): Sample Design Manual, Version 2.0. Atlanta, GA: Centers for Disease Control and Prevention, 2010.

Table 1 Model-based estimates of current tobacco use by sex, province in Thailand

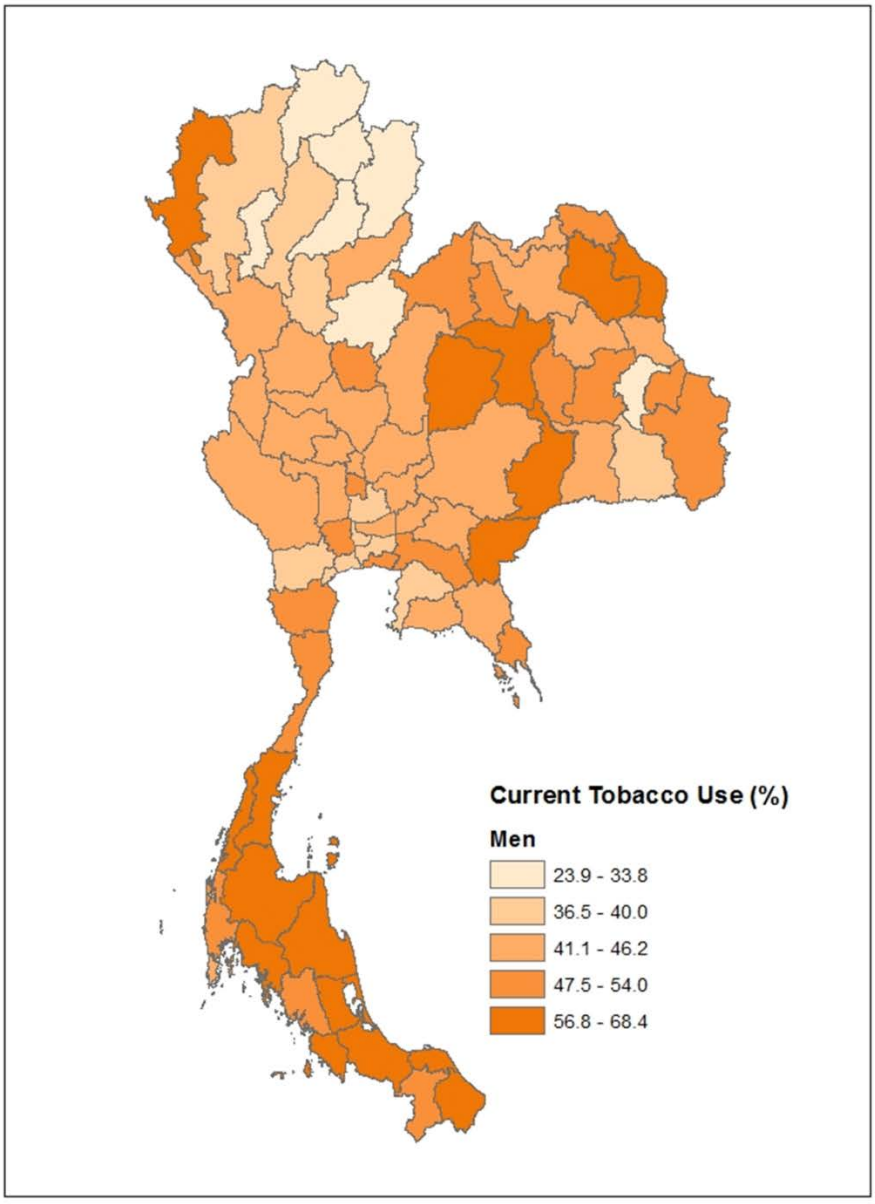
SEX	N	Min	Q1	Median	Q2	Max	Mean
Male	77	24.4	40.6	46.5	54.2	68.2	47.6
Female	77	1.41	4.31	6.46	9.49	24.7	7.67

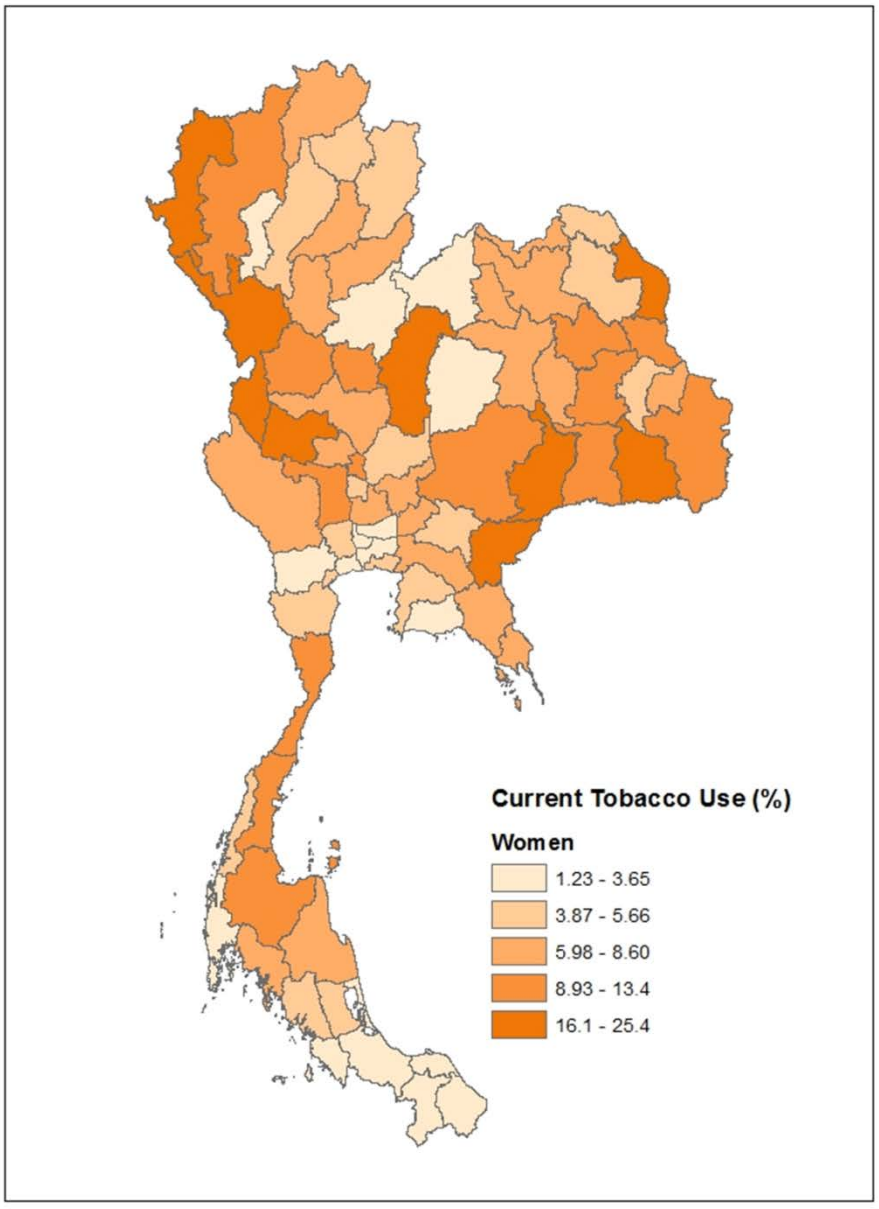
Table 2 The correlation between model-based and direct survey estimates of current tobacco use by sex

Weighted	Sex	Method	N	Mean	Std	Minimum	Maximum	Rho*
YES	Male	model	65	47.0	12.3	17.8	70.3	0.969
		survey	65	47.7	10.3	24.4	68.2	
	Female	model	67	7.71	5.61	0.42	25.4	0.945
		survey	67	7.73	5.11	1.41	24.7	
NO	Male	model	65	47.0	12.3	17.8	70.3	0.773
		survey	65	46.8	8.2	30.3	68.3	
	Female	model	67	7.71	5.61	0.42	25.4	0.906
		survey	67	7.56	4.12	1.96	20.0	

*limited to those provinces with at least 30 respondents.

Figure Prevalence of current tobacco use for men and women at provincial-level in Thailand 2011




Current Tobacco Use (%)
Women
1.23 - 3.65
3.87 - 5.66
5.98 - 8.60
8.93 - 13.4
16.1 - 25.4