A comparison of methods for missing covariates in a meta-regression using data from a systematic review on oral epithelial dysplasia

Yeonggyeong Jang[1,2], Areum Han[1,2], Hee-Kyung Park[3,4], Seokyung Hahn[5,6,7]

[1]Interdisciplinary Program of Medical Informatics, Seoul National University College of Medicine, Seoul, Korea
[2] Integrated Major in Innovative Medical Science, Seoul National University Graduate School, Seoul, Korea
[3]Department of Oral Medicine, Seoul National University Dental Hospital, Seoul, Korea
[4]Department of Oral Medicine and Oral Diagnosis, School of Dentistry and Dental Research Institute, Seoul National University, Seoul, Korea
[5]Department of Human Systems Medicine, Seoul National University College of Medicine, Seoul, Korea
[6]Institute of Health Policy and Management, Medical Research Center, Seoul National University, Seoul, Korea
[7]Medical Big Data Research Center, Medical Research Center, Seoul National University, Seoul, Korea

ISPOR
Improving healthcare decisions

MSR1

Background

- Missing data is a common issue in meta-analyses. This problem can be more critical in meta-regression, which usually involves a number of covariates and therefore encounters more studies being omitted due to missing covariates.

Objectives

- The aim of this study was to explore applicability of methods for handling missing covariates in meta-regression analyses.

Methods

- We extracted data from 54 studies included in a systematic review evaluating malignant transformation rate of oral epithelial dysplasia. The data contained 10 study-level covariates.
- A plot was created to visualize and to identify the distribution and patterns of the missing covariates.
- A conventional multiple random-effects meta-regression model was constructed through manual backward elimination based on complete cases.
- Three methods were employed to analyze the full dataset: A Bayesian random-effects model; multiple imputation using MICE (multivariate imputation by chained equations); and a full information maximum likelihood (FIML) model.
- We compared the regression coefficient and p-value from each method.

Results

- Based on the missing pattern plot, 42.6% of the studies had missing data for all four covariates: alcohol consumption, gender, smoking status, and dysplasia site on tongue or floor of the mouth (figure 1).
- Conventional complete case analysis only considered 10 (18.5%) of the 54 studies. After manual backward elimination, the model included severity, dysplasia site, smoking status and follow-up as covariates (table 1). The coefficients of all the covariates indicated a positive relationship to the malignant transformation rate from oral epithelial dysplasia.
- A total of 26 studies had no more than three missing covariates.
- The coefficients estimated by Bayesian and FIML models were very similar to those from the complete case analysis; however, two covariates (smoking, follow-up duration) were no longer statistically significant. When the MICE was employed, the size of all the coefficients decreased by half at the most, while one covariate (smoking) lost its statistical significance.
- Smoking status was the covariate that lost statistical significance after consideration of missing data. Smoking was the covariate that had missing values in more than half of the studies in the dataset.

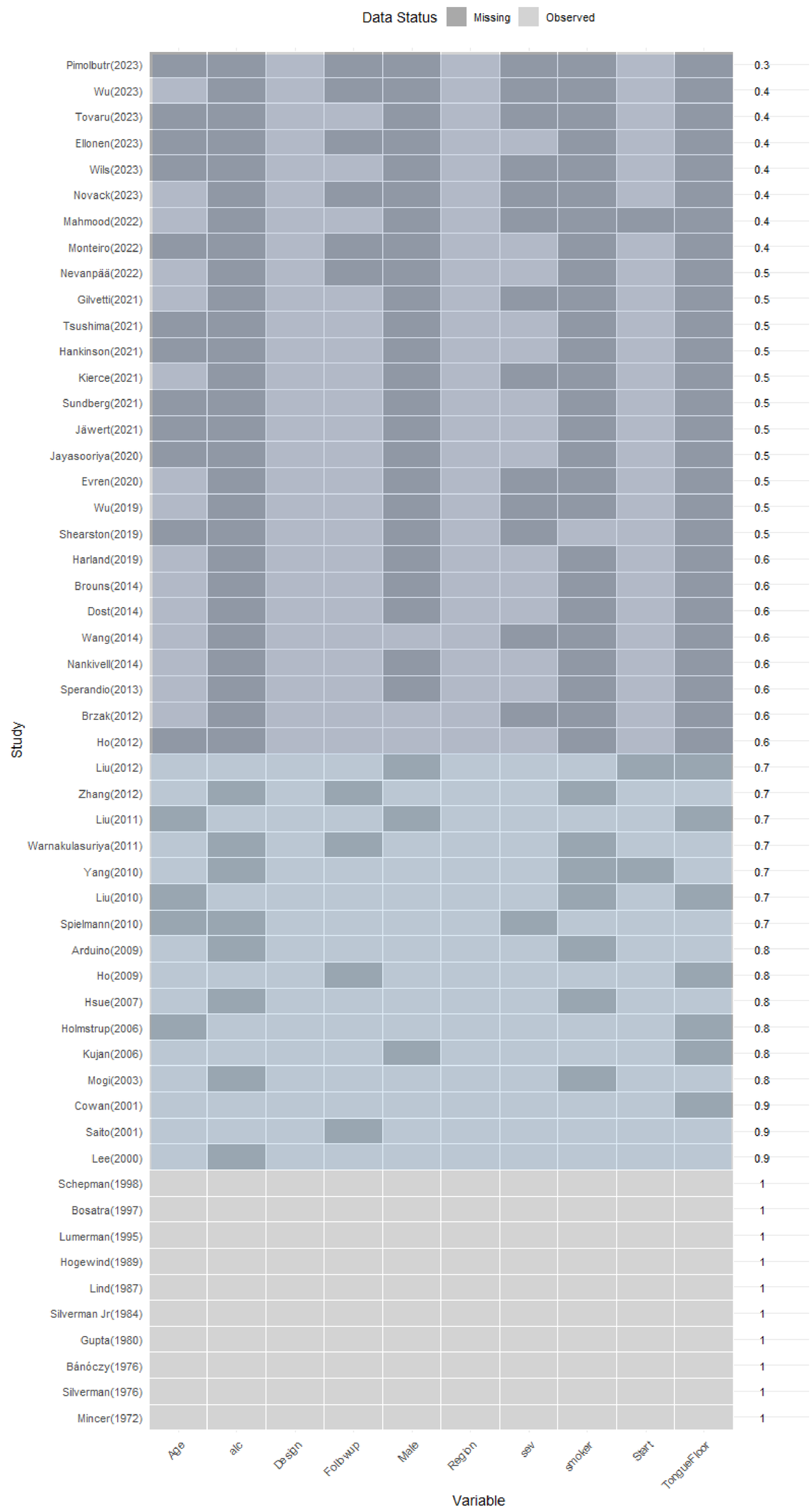

Figure 1. Missing pattern plot

	Complete case analysis (studies analyzed = 11)		Bayesian		Multiple imputation using MICE (m= 500)		Multiple imputation using MICE (m= 500) (in 26 studies)		FIML	
Covariate	Coefficient estimate	p-value	Coefficient estimate	CI (Credible interval)	Coefficient estimate	p-value	Coefficient estimate	p-value	Coefficient estimate	p-value
Proportion of severe dysplasia	3.860	<.001	3.860	0.335–7.288	1.876	<.001	1.077	0.019	3.839	0.023
Proportion of site on tongue or floor	4.062	<.001	4.063	1.116–6.978	3.302	<.001	3.301	<.001	4.089	0.006
Proportion of ever smoker	1.034	0.023	1.042	-0.793–3.280	0.416	0.287	0.243	0.543	1.162	0.248
Mean follow-up duration	0.215	0.006	0.215	-0.028–0.532	0.189	<.001	0.282	<.001	0.244	0.100

Table 1. Results of a random-effects multiple meta-regression

Conclusions

- We demonstrated that multiple meta-regression can be based on a very small proportion of data due to missing covariates, which can lead to the effects of clinically significant covariates being over- or underestimated. Therefore, it is important to use multiple approaches to handle missing data and explore how sensitive the conclusions from the original analysis are.

Conflict of Interest

- This research was supported by a grant from the National Research Foundation of Korea (Grants No.: 2022R1A2C101043711 and No.: RS-2025-00559766).